Attorney Docket No.: 2307o-099910US

PATENT APPLICATION

NUCLEIC ACIDS THAT CONTROL PLANT DEVELOPMENT

Inventor(s):

Robert L. Fischer, a citizen of the United States
residing at: 1423 Scott Street, El Cerrito, CA 94530

Yeonhee Choi, a citizen of South Korea
residing at: 4 Anchor Dr. #442, Emeryville, CA 94608

Mike Hannon, a citizen of the United States
residing at: 309 Pearl Drive, Livermore, CA 94550

Jack Kishiro Okamuro, a citizen of the United States
residing at: 307 Ocho Rios Way, Oak Park, CA 91377

Tatiana Valerievna Tatarinova, a citizen of Russia
residing at 836 North Beaudry Avenue Apt. 2, Los Angeles, CA 90012

Assignee:

THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
1111 Franklin Street, Fifth Floor
Oakland, CA 94607-5200

Entity:

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 415-576-0200

NUCLEIC ACIDS THAT CONTROL PLANT DEVELOPMENT

CROSS-REFERENCES TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Patent Application No. 09/553,690, filed April 21, 2000, the contents of which are incorporated by reference.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This invention was made with Government support under Grant No. 97-35304-4941, awarded by the United States Department of Agriculture. The government has certain rights in this invention.

FIELD OF THE INVENTION

This invention is directed to plant genetic engineering. It relates to, for example, modulating seed (and in particular endosperm, embryo and seed coat) development, flowering time, chromosomal DNA methylation and modulating transcription in plants.

BACKGROUND OF THE INVENTION

A fundamental problem in biology is to understand how seed development. In flowering plants, the ovule generates the female gametophyte, which is composed of egg, central, synergid and antipodal cells (Reiser, *et al., Plant Cell*, 1291-1301 (1993)). All are haploid except the central cell which contains two daughter nuclei that fuse prior to fertilization. One sperm nucleus fertilizes the egg to form the zygote, whereas another sperm nucleus fuses with the diploid central cell nucleus to form the triploid endosperm nucleus (van Went, *et al.*, *Embryology of Angiosperms*, pp. 273-318 (1984)). The two fertilization products undergo distinct patterns of development. In Arabidopsis, the embryo passes through a series of stages that have been defined morphologically as preglobular, globular, heart, cotyledon and maturation (Goldberg, R. B., *et al.*, *Science* (1994) 266: 605-614; Mansfield, S. G., *et al.*, *Arabidopsis: An Atlas of Morphology and Development*, pp. 367-383 (1994)). The primary endosperm nucleus undergoes a series of mitotic divisions to produce nuclei that migrate into the expanding central cell (Mansfield, S. G., *et al.*, *Arab Inf Serv* 27: 53-64 (1990); Webb, M. C., *et al.*, *Planta* 184:187-195 (1991)). Cytokinesis sequesters endosperm cytoplasm and nuclei

into discrete cells (Mansfield, S. G., *et al.*, *Arab Inf Serv* 27:65-72 (1990)) that produce storage proteins, starch, and lipids which support embryo growth (Lopes, M. A. *et al.*, *Plant Cell* 5:1383-1399 (1993)). Fertilization also activates development of the integument cell layers of the ovule that become the seed coat, and induces the ovary to grow and form the fruit, or silique, in Arabidopsis.

Of particular interest are recent discoveries of genes that control seed, and in particular endosperm, development. For instance, *MEDEA* (*MEA*) (also known as *FIE1* (see, e.g., copending U.S. patent application 09/071,838) and F644 (see, e.g., Kiyosue T, *et al.* (1999) *Proc Natl Acad Sci U S A* 96(7):4186-91) encodes an Arabidopsis SET domain polycomb protein that appears to play a role in endosperm development. Inheritance of a maternal loss-of-function mea allele results in embryo abortion and prolonged endosperm production, irrespective of the genotype of the paternal allele. Thus, only the maternal wild-type MEA allele is required for proper embryo, endosperm, and seed coat development (Kinoshita T, *et al.* (1999) *Plant Cell* 10:1945-52). These results reveal functions for plant polycomb proteins in the suppression of central cell proliferation and endosperm development (Kiyosue T, *et al. supra*).

Another gene product that controls seed development is FIE, also known as FIE3 (see, e.g., copending U.S. patent application 09/071,838). The FIE protein is a homolog of the WD motif-containing Polycomb proteins from Drosophila and mammals (Ohad, N. *et al. Plant Cell* 11(3):407-16 (1999)). In Drosophila, these proteins function as repressors of homeotic genes. Loss of function mutations in the *FIE* gene result in endosperm phenotypes that are identical to *medea* loss of function mutations. A female gametophyte with a loss-of-function allele of *fie* undergoes replication of the central cell nucleus and initiates endosperm development without fertilization. These results suggest that the FIE Polycomb protein functions to suppress a critical aspect of early plant reproduction, namely, endosperm development, until fertilization occurs. Moreover, hypomethylation of *fie* mutants leads to the development of differentiated endosperm. Vinkenoog *et al.*, *Plant Cell* 12:2271-2282 (2000).

Control of the expression of genes that control egg and central cell differentiation, or those that control reproductive development, i.e. embryo, endosperm and seed coat, is useful in the production of plants with a range of desired traits. These and other advantages are provided by the present application.

SUMMARY OF THE INVENTION

This invention provides isolated nucleic acids comprising a polynucleotide sequence, or its complement, encoding a DMT polypeptide comprising an amino acid sequence with at least 70% sequence identity to at least one of the following consensus sequences:

DMT Domain A

KV<1>(I,l)D(D,p)(E,v)T<3>W<1>(L,v)L(M,l)(E,d)<0-2>D(K,e)<1>(K,t)<1>(K,a)(W,k)(W,l)<1>(E,k)ER<2>F<1>(G,t)R<1>(D,n)(S,l)FI(A,n)RM(H,r)<1>(V,l)QG(D,n)R<1>F<1>(P,q)WKGSVVDSV(I,v)GVFLTQN(V,t)D(H,y)(L,s)SS(S,n)A(F,y)M<1>(L,v)A(A,s)<1>FP

DMT Domain B

W(D,n)<1>(L,f)R<5>E<3-6>D(S,t)<1>(D,n)(Y,w)<3>R<10>I<2>RG(M,q)(N,f)<2>L(A,s)<1>RI<2-12>FL<3>V<2>(H,n)G<1>IDLEWLR<2>(P,d)(P,s)(D,h)<1>(A,v)K<1>(Y,f)LL(S,e)(I,f)<1>G(L,i)GLKS(V,a)ECVRLL<1>L(H,k)<2>AFPVDTNVGRI(A,c)VR(M,l)G(W,l)VPL(Q,e)PLP<2>(L,v)Q(L,m)H(L,q)L(E,f)<1>YP<1>(L,m)(E,d)(S,n)(I,v)QK(F,y)LWPRLCKL(D,p)Q<1>TLYELHY(Q,h)(L,m)ITFGK<0-2>FCTK<2>PNCNACPM(R,k)<0-2>EC(R,k)(H,y)(F,y)(A,s)SA<1>(A,v)<0-10>S(A,s)(R,k)<1>(A,l)L(P,e)<1>(P,t)

DMT Domain C.

P(I,l)(I,v)E(E,f)P<1>(S,t)P<2-5>E<0-15>(D,a)IE(D,e)<4-23>(I,v)P<1>I<1>(L,f)(N,d)<8-17>(S,a)<1>(A,d)LV<8>(I,l)P<2-5>(K,r)(L,m)K<4>LRTEH<1>V(Y,f)(E,v)LPD<1>H<1>(L,i)L(E,k)<1>(D,e)D(P,i)<2>YLL(A,s)IW(T,q)P(G,d)(E,g)<6-8>(P,s)<3>C<6-10>(M,l)C<4>C<2>C<3>(R,k)E<5>(V,f)RGT(L,i)L<0-22>(L,v)FADH<1>(S,t)(S,r)<2>PI<3>(R,t)<3>(W,k)<1>L<1>(R,k)R<4>G(T,s)(S,t)<2>(S,t)I(F,c)(R,k)(G,l)L<1>(T,v)<2>I<2>(C,n)F(W,q)<1>G(F,y)(V,l)C(V,l)R<1>F(E,d)<3>(R,g)<1>P(R,k)<1>L<2>(R,h)LH<2>(A,v)SK

In some embodiments, the nucleic acids of the invention do not encode a polypeptide at least 40% identical to SEQ ID NO:2, or alternatively at least 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% to SEQ ID NO:2. In some embodiments, the DMT polypeptide comprises an amino acid sequence 100% identical to the above-listed consensus sequences.

In some embodiments, the DMT polypeptides ar at least 45%,50%,55%,60%,65%,70%,75%,80%,85%,90%,95%,97%,98%,99% or 100% identical to DMT domains A, B and/or C.

In one aspect, the invention provides DMt polypeptides capable of exhibiting at least one of the following biological activities:

(a) glycosylase activity;

(b) demethylation of polynucleotides;

(c) DNA repair;

(d) wherein expression of the polypeptide in a plant modulates organ identity;

(e) wherein expression of the polypeptide in a plant modulates organ number;

(f) wherein expression of the polypeptide in a plant modulate meristem stem and/or activity;

(g) wherein enhanced expression of the polypeptide in a plant results in a delay in flowering time;

(h) wherein introduction of the polypeptide into a cell results in modulation of methylation of chromosomal DNA in the cell;

(i) wherein reduction of expression of the polypeptide in a plant results in modulation of endosperm development;

(j) wherein expression of the polypeptide in an Arabidopsis leaf results in modulation of expression of the MEDEA gene.

In some aspects, the polypeptide comprises either a

(i) basic region;

(ii) nuclear localization signal;

(iii) leucine zipper;

(iv) helix-hairpin-helix structure;

(v) glycine-proline rich loop with a terminal aspartic acid or

(vi) helix that is capable of binding DNA.

In one aspect, the invention provides methods of modulating in a plant one or more of the following:

(a) DNA repair;

(b) wherein expression of the polypeptide in a plant modulates organ identity;

(c) wherein expression of the polypeptide in a plant modulates organ number;

(d) wherein expression of the polypeptide in a plant modulate meristem stem and/or activity;

(e) wherein enhanced expression of the polypeptide in a plant results in a delay in flowering time;

(f) wherein introduction of the polypeptide into a cell results in modulation of methylation of chromosomal DNA in the cell;

(g) wherein reduction of expression of the polypeptide in a plant results in modulation of endosperm development;

(h) wherein expression of the polypeptide in an Arabidopsis leaf results in expression of the MEDEA gene,

wherein the method comprises:,

(a) introducing into a plant cell a nucleic acid of claim 1; and

(b) generating conditions where the plant cell can transcribe the nucleic acid described above.

In some embodiments, the polypeptides comprise between 1500 and 2000 amino acids. In some aspects, the polypeptide has glycosylase activity. In some embodiments, introduction of the nucleic acid into a cell results in modulation of methylation of chromosomal DNA in the cell. In some embodiments, enhanced expression of the nucleic acids of the invention into a plant results in a delay in flowering time. In some embodiments, reduction of expression of a DMT polypeptide in a plant results in enhanced endosperm development. In addition, in some embodiments, expression of the nucleic acid of the invention in an *Arabidopsis* leaf results in expression of the *MEDEA* gene.

This invention provides isolated nucleic acids comprising a polynucleotide sequence, or its complement, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C. For instance, the nucleic acid can encode the DMT polypeptide displayed in SEQ ID NO:2. In one aspect, the polynucleotide sequence comprises SEQ ID NO:5 or SEQ ID NO:1. In some aspects of the invention, the nucleic acid further comprises a promoter operably linked to the polynucleotide. In some embodiments, the promoter is constitutive. In other embodiments, the promoter is from a DMT gene. For example, the promoter can comprise a polynucleotide at least 70% identical to SEQ ID NO:3. In some aspects, the promoter comprises SEQ ID NO:3. In some aspects of this invention, the promoter further comprises a polynucleotide at least

70% identical to SEQ ID NO:4. For example, in some aspects the promoter comprises SEQ ID NO:4. In some aspects, the polynucleotide sequence is linked to the promoter in an antisense orientation.

The invention also provides an isolated nucleic acid molecule comprising a polynucleotide sequence exhibiting at least 60% sequence identity to SEQ ID NO:1.

The invention also provides an expression cassette comprising a promoter operably linked to a heterologous polynucleotide sequence, or complement thereof, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2. For instance, the nucleic acid can encode the DMT polypeptide displayed in SEQ ID NO:2. In some aspects, the polynucleotide sequence comprises SEQ ID NO:5 or SEQ ID NO:1. In some aspects of the invention, the nucleic acid further comprises a promoter operably linked to the polynucleotide. In some embodiments, the promoter is constitutive. In other embodiments, the promoter is from a DMT gene. For example, the promoter can comprise a polynucleotide at least 70% identical to SEQ ID NO:3. In some aspects, the promoter comprises SEQ ID NO:3. In some aspects of this invention, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:4. For example, in some aspects the promoter comprises SEQ ID NO:4. In some aspects, the polynucleotide sequence is linked to the promoter in an antisense orientation.

The invention also provides an expression cassette for the expression of a heterologous polynucleotide in a plant cell. In some aspects, the expression cassette comprises a promoter polynucleotide at least 70% identical to SEQ ID NO:3 that is operably linked to a heterologous polynucleotide. In some aspects, the promoter comprises SEQ ID NO:3. In some aspects, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:4. For instance, in some embodiments, the promoter comprises SEQ ID NO:4. In some aspects, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:6. In some aspects, the promoter comprises SEQ ID NO:6.

The present invention also provides a host cell comprising an exogenous polynucleotide sequence comprising a polynucleotide sequence, or complement thereof, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C. In some aspects of the invention, the nucleic acid further comprises a promoter operably linked to the polynucleotide sequence. In some aspects, the promoter is constitutive. In some aspects, the promoter comprises a polynucleotide at least 70% identical to SEQ ID

NO:3. The promoter, for instance, can comprise SEQ ID NO:3. In some aspects, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:4. For instance, in some embodiments, the promoter comprises SEQ ID NO:4. In some aspects, the promoter is operably linked to the exogenous polynucleotide sequence in an antisense orientation.

The present invention also provides an isolated polypeptide comprising an amino acid sequence at least 60% identical to SEQ ID NO:2 or an amino acid sequence at least 70% sequence identical to at least one of DMT domain A, B, or C and capable of exhibiting at least one biological activity of the polypeptide displayed in SEQ ID NO:2, or fragment thereof. The present invention also provides for an antibody capable of binding such polypeptides.

The present invention also provides a method of introducing an isolated nucleic acid into a host cell comprising, (a) providing an isolated nucleic acid or its complement, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C and (b) contacting the nucleic acid with the host cell under conditions that permit insertion of the nucleic acid into the host cell.

The present invention also provides a method of modulating transcription, comprising introducing into a host cell an expression cassette comprising a promoter operably linked to a heterologous DMT polynucleotide, the heterologous DMT polynucleotide encoding a DMT polypeptide at least 60% identical to SEQ ID NO:2 or at least 70% sequence identical to at least one of DMT domain A, B, or C, and detecting a host cell with modulated transcription. In some aspects of the invention, the heterologous DMT polynucleotide encodes SEQ ID NO:2. In some aspect, the polynucleotide sequence comprises SEQ ID NO:5 or SEQ ID NO:1. In some aspects, the expression cassette is introduced into a host cell by *Agrobacterium*. In some aspects, the expression cassette is introduced by a sexual cross. In some aspects of the method of the invention, modulating transcription results in the modulation of endosperm development in a plant. In some aspects, endosperm development is enhanced. In other aspects, endosperm development is decreased. In some aspects of the methods of the invention, the promoter is operably linked to the DMT polynucleotide in an antisense orientation.

The present invention also provides a method of detecting a nucleic acid in a sample, comprising (a) providing an isolated nucleic acid molecule comprising a polynucleotide sequence, or its complement, encoding a DMT polypeptide exhibiting at

least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C, (b) contacting the isolated nucleic acid molecule with a sample under conditions that permit a comparison of the sequence of the isolated nucleic acid molecule with the sequence of DNA in the sample, and (c) analyzing the result of the comparison. In some aspects of the method, the isolated nucleic acid molecule and the sample are contacted under conditions that permit the formation of a duplex between complementary nucleic acid sequences.

The present invention also provides a transgenic plant cell or transgenic plant comprising a polynucleotide sequence, or its complement, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C. For instance, the nucleic acid can encode the DMT polypeptide displayed in SEQ ID NO:2. In one aspect, the polynucleotide sequence comprises SEQ ID NO:5 or SEQ ID NO:1. In some aspects of the invention, the nucleic acid further comprises a promoter operably linked to the polynucleotide. In some embodiments, the promoter is constitutive. In other embodiments, the promoter comprises a polynucleotide at least 70% identical to SEQ ID NO:3. In some aspects, the promoter comprises SEQ ID NO:3. In some aspects of this invention, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:4. For example, in some aspects the promoter comprises SEQ ID NO:4. In some aspects, the polynucleotide sequence is linked to the promoter in an antisense orientation. The present invention also provides a plant that is regenerated from a plant cell as described above.

The present invention also provides an expression cassette for the expression of a heterologous polynucleotide in a plant cell, wherein the expression cassette comprises a promoter at least 70% identical to SEQ ID NO:3 and the promoter is operably linked to a heterologous polynucleotide. In some embodiments, the promoter comprises a polynucleotide at least 70% identical to SEQ ID NO:4 and/or SEQ ID NO:6. In some embodiments, the promoter specifically directs expression of the heterologous polynucleotide in a female gametophyte when the expression cassette is introduced into a plant.

DEFINITIONS

The phrase "nucleic acid sequence" refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. It

includes chromosomal DNA, self-replicating plasmids, infectious polymers of DNA or RNA and DNA or RNA that performs a primarily structural role.

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of an operably linked nucleic acid. As used herein, a "plant promoter" is a promoter that functions in plants. Promoters include necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

The term "plant" includes whole plants, plant organs (e.g., leaves, stems, flowers, roots, etc.), seeds and plant cells and progeny of same. The class of plants which can be used in the method of the invention is generally as broad as the class of flowering plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), as well as gymnosperms. It includes plants of a variety of ploidy levels, including polyploid, diploid, haploid and hemizygous.

A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants.

A polynucleotide "exogenous to" an individual plant is a polynucleotide which is introduced into the plant, or a predecessor generation of the plant, by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include Agrobacterium-mediated transformation, biolistic methods, electroporation, in planta techniques, and the like. "Exogenous," as referred to within, is any polynucleotide, polypeptide or protein sequence, whether chimeric or not, that is initially or subsequently introduced into the genome of an individual host cell or the

organism regenerated from said host cell by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include Agrobacterium-mediated transformation (of dicots - e.g. Salomon et al. EMBO J. 3:141 (1984); Herrera-Estrella et al. EMBO J. 2:987 (1983); of monocots, representative papers are those by Escudero et al., Plant J. 10:355 (1996), Ishida et al., Nature Biotechnology 14:745 (1996), May et al., Bio/Technology 13:486 (1995)), biolistic methods (Armaleo et al., Current Genetics 17:97 1990)), electroporation, in planta techniques, and the like. Such a plant containing the exogenous nucleic acid is referred to here as a T0 for the primary transgenic plant and T1 for the first generation. The term "exogenous" as used herein is also intended to encompass inserting a naturally found element into a non-naturally found location.

The phrase "host cell" refers to a cell from any organism. Preferred host cells are derived from plants, bacteria, yeast, fungi, insects or other animals, including humans. Methods for introducing polynucleotide sequences into various types of host cells are well known in the art.

The "biological activity of a polypeptide" refers to any molecular activity or phenotype that is caused by the polypeptide. For example, the ability to transfer a phosphate to a substrate or the ability to bind a specific DNA sequence is a biological activity. One biological activity of DMT is glycosylase activity, i.e., cleavage of the nucleotide base from the nucleotide sugar). Another biological activity of DMT is to demethylate nucleotides (e.g., DMT has 5'-methylcytosine glycosylase activity). In addition, DMT has the ability to modulate endosperm production, as described herein, and to modulate flowering time in plants. For example, when DMT expression or DMT activity is increased in a plant, the flowering time of the plant is delayed. Moreover, expression of a DMT polypeptide in a plant tissue (e.g., a leaf) that does not typically express the *MEDEA* gene (Grossniklaus U, *et al.*, *Science* 280(5362):446-50 (1998)) results in the expression of *MEDEA*.

Additional biological activities of DMT polypeptides include: nuclear localization (e.g., as localized by amino acids 43-78 of SEQ ID NO:2); the ability to modulate plant organ size and/or number; the ability to modulate meristem size and/or activity; and to perform DNA repair, including nucleotide methylation or demethylation and/orrepair and/or removal of mis-matched nucleotides from DNA.

An "expression cassette" refers to a nucleic acid construct, which when introduced into a host cell, results in transcription and/or translation of an RNA or

polypeptide, respectively. Antisense or sense constructs that are not or cannot be translated are expressly included by this definition.

Sub a2> A "*DMT* nucleic acid" or "*DMT* polynucleotide sequence" of the invention is a subsequence or full length polynucleotide sequence of a gene which encodes a polypeptide involved in control of reproductive development and which, when the maternal allele is mutated or when DMT activity is reduced or eliminated in a maternal tissue or plant, allows for increased production of the endosperm and/or abortion of the embryo. In addition, overexpression of DMT in plants results in delayed time to flowering. Moreover, DMT is necessary and sufficient for expression of *MEDEA* in a plant cell. An exemplary nucleic acid of the invention is the *Arabidopsis DMT* sequence (SEQ ID NO:1). Additional *DMT* nucleic acid sequences from a variety of plant species are also provided (e.g., SEQ ID NOs: 7-70). *DMT* polynucleotides are defined by their ability to hybridize under defined conditions to the exemplified nucleic acids or PCR products derived from them. A DMT polynucleotide is typically at least about 30-40 nucleotides to about 7000, usually less than about 10,000 nucleotides in length. More preferably, DMT polynucleotides contain a coding sequence of from about 100 to about 5500 nucleotides, often from about 500 to about 3600 nucleotides in length. A DMT polypeptide is typically at least 500 amino acids, typically at least 1000 amino acids, more typically at least 1500 amino acids. In some embodiments, a DMT polypeptide comprises fewer than 2000 amino acids, more typically fewer than 3000 amino acid and still more typically fewer than 5000 or 7500 amino acid in length.

Sub a3> As described below, *DMT* nucleic acid sequences encode polypeptides with substantial identity to at least one of following the consensus sequences:

DMT Domain A

KV<1>(I,l)D(D,p)(E,v)T<3>W<1>(L,v)L(M,l)(E,d)<0-2>D(K,e)<1>(K,t)<1>(K,a)(W,k)(W,l)<1>(E,k)ER<2>F<1>(G,t)R<1>(D,n)(S,l)FI(A,n)RM(H,r)<1>(V,l)QG(D,n)R<1>F<1>(P,q)WKGSVVDSV(I,v)GVFLTQN(V,t)D(H,y)(L,s)SS(S,n)A(F,y)M<1>(L,v)A(A,s)<1>FP

DMT Domain B

W(D,n)<1>(L,f)R<5>E<3-6>D(S,t)<1>(D,n)(Y,w)<3>R<10>I<2>RG(M,q)(N,f)<2>L(A,s)<1>RI<2-12>FL<3>V<2>(H,n)G<1>IDLEWLR<2>(P,d)(P,s)(D,h)<1>(A,v)K<1>(Y,f)LL(S,e)(I,f)<1>G(L,i)GLKS(V,a)ECVRLL<1>L(H,k)<2>AFPVDTNVGRI(A,c)VR(M,l)G(W,l)VPL(Q,e)PLP<2>(L,v)Q(L,m)H(L,q)L(E,f)<1>YP<1>(L,m)(E,d)(S,n)(I,v)QK(F,y)LWPRLCKL(D,p)Q<1>TLYELHY(Q,h

)(L,m)ITFGK<0-2>FCTK<2>PNCNACPM(R,k)<0-2>EC(R,k)(H,y)(F,y)(A,s)SA<1>(A,v)<0-10>S(A,s)(R,k)<1>(A,l)L(P,e)<1>(P,t)

DMT Domain C.

P(I,l)(I,v)E(E,f)P<1>(S,t)P<2-5>E<0-15>(D,a)IE(D,e)<4-23>(I,v)P<1>I<1>(L,f)(N,d)<8-17>(S,a)<1>(A,d)LV<8>(I,l)P<2-5>(K,r)(L,m)K<4>LRTEH<1>V(Y,f)(E,v)LPD<1>H<1>(L,i)L(E,k)<1>(D,e)D(P,i)<2>YLL(A,s)IW(T,q)P(G,d)(E,g)<6-8>(P,s)<3>C<6-10>(M,l)C<4>C<2>C<3>(R,k)E<5>(V,f)RGT(L,i)L<0-22>(L,v)FADH<1>(S,t)(S,r)<2>PI<3>(R,t)<3>(W,k)<1>L<1>(R,k)R<4>G(T,s)(S,t)<2>(S,t)I(F,c)(R,k)(G,l)L<1>(T,v)<2>I<2>(C,n)F(W,q)<1>G(F,y)(V,l)C(V,l)R<1>F(E,d)<3>(R,g)<1>P(R,k)<1>L<2>(R,h)LH<2>(A,v)SK

In addition, the following cansensus sequence spanning all three domains were identified:

<9-14>(T,q)(A,i)(S,k)(I,l)<3>(A,r)(S,k)<1>(G,m)<2>(S,r)(P,k)<2>(K,f)<2>(E,l)K<0-1>K<0-3>(P,r)<2>(P,r)<1>(K,r)(K,r)(G,d)(R,k)<1>(G,v)<1>(K,g)<3-5>(P,s)(P,k)<3>(S,n)<1>(I,l)<0-2>(Q,d)<9>(P,q)<4>(K,a)(P,s)<14-16>(P,a)<4>L<0-10>D<1>(I,l)<0-4>(L,n)<12-46>(K,d)<2-7>(P,a)KV<1>(I,l)D(D,p)(E,v)T<3>W<1>(L,v)L(M,l)(E,d)<0-2>D(K,e)<1>(K,t)<1>(K,a)(W,k)(W,l)<1>(E,k)ER<2>F<1>(G,t)R<1>(D,n)(S,l)FI(A,n)RM(H,r)<1>(V,l)QG(D,n)R<1>F<1>(P,q)WKGSVVDSV(I,v)GVFLTQN(V,t)D(H,y)(L,s)SS(S,n)A(F,y)M<1>(L,v)A(A,s)<1>PP<0-16>(P,v)<6-15>(S,h)<3>(E,d)<10-24>(S,t)<1>(S,e)<6>(K,n)<8-55>(E,i)<8-9>(I,v)<1>(N,s)<1-4>(E,d)<1>(E,s)<4>(Q,l)<0-11>(D,h)<1>(F,m)<5>(Q,n)<0-3>(G,e)<2>(G,d)S<1>(K,d)<7-11>(T,m)<2>(V,l)<3>(S,q)<6-10>(S,e)<2-3>(S,v)<19-25>(T,s)<16-28>(R,s)<2-6>(T,p)<5>(P,k)<10>(Q,e)<4>(D,s)<1-4>(S,r)<5>(D,p)<3>(N,d)<3>(P,y)<2>(F,s)<1>(R,k)<1>(G,s)<1>(S,a)(V,r)(P,e)<3>(T,s)<3-6>(I,l)<3>(P,e)<1>E<3-5>(L,q)<1>(G,c)<1>(S,h)(S,n)<1>(V,q)<1>(E,d)<3>T(Q,e)<1-2>(N,g)<3>(E,n)<20-30>(N,a)(P,g)<1-6>(S,l)<25-46>(Q,d)W(D,n)<1>(L,f)R<5>E<3-6>D(S,t)<1>(D,n)(Y,w)<3>R<10>I<2>RG(M,q)(N,f)<2>L(A,s)<1>RI<2-12>FL<3>V<2>(H,n)G<1>IDLEWLR<2>(P,d)(P,s)(D,h)<1>(A,v)K<1>(Y,f)LL(S,e)(I,f)<1>G(L,i)GLKS(V,a)ECVRLL<1>L(H,k)<2>AFPVDTNVGRI(A,c)VR(M,l)G(W,l)VPL(Q,e)PLP<2>(L,v)Q(L,m)H(L,q)L(E,f)<1>YP<1>(L,m)(E,d)(S,n)(I,v)QK(F,y)LWPRLCKL(D,p)Q<1>TLYELHY(Q,h)(L,m)ITFGK<0-2>FCTK<2>PNCNACPM(R,k)<0-2>EC(R,k)(H,y)(F,y)(A,s)SA<1>(A,v)<0-10>S(A,s)(R,k)<1>(A,l)L(P,e)<1>(P,t)(E,q)<7-16>P(I,l)(I,v)E(E,f)P<1>(S,t)P<2-5>E<0-15>(D,a)IE(D,e)<4-23>(I,v)P<1>I<1>(L,f)(N,d)<8-17>(S,a)<1>(A,d)LV<8>(I,l)P<2-5>(K,r)(L,m)K<4>LRTEH<1>V(Y,f)(E,v)LPD<1>H<1>(L,i)L(E,k)<1>(D,e)D(P,i)<2>YLL(A,s)IW(T,q)P(G,d)(E,g)<6-8>(P,s)<3>C<6-10>(M,l)C<4>C<2>C<3>(R,k)E<5>(V,f)RGT(L,i)L<0-22>(L,v)FADH<1>(S,t)(S,r)<2>PI<3>(R,t)<3>(W,k)<1>L<1>(R,k)R<4>G(T,s)(S,t)<2>(S,t

)I(F,c)(R,k)(G,l)L<1>(T,v)<2>I<2>(C,n)F(W,q)<1>G(F,y)(V,l)C(V,l)R<1>F(E,d)<3>(R,g)<1>P(R,k)<1>L<2>(R,k)LH<2>(A,v)SK

DMT domain A corresponds to amino acid positions 697 through 796 of SEQ ID NO:2. DMT domain B corresponds to amino acid positions 1192 through 1404 of SEQ ID NO:2. DMT domain C corresponds to amino acid positions 1452 through 1722 of SEQ ID NO:2. The consensus sequence provides amino acid sequences by position using single letter amino acid abbreviations. Numbers in carrots ("<" or ">") refer to amino acid positions where there is no consensus and which therefore, can be any amino acid. Amino acid abbreviations in parentheses indicate alternative amino acids at the same position. Capitalized letters refer to predominant consensus amino acids and lower case letters refer to amino acids that are commonly found in DMT sequences, but are not predominant. Thus, it is a simple matter to identify whether any particular nucleic acid sequence is a DMT nucleic acid and/or encodes a DMT polypeptide.

The structure of full-length DMT polypeptides comprises the following domains and regions. These regions are generally described with reference to SEQ ID NO:2. First, as described above, domain B DMT polypeptides can comprise a bipartite nuclear localization signal (e.g., amino acid positions 43-60 and 61-78 in SEQ ID NO:2) comprised of basic amino acids. Amino acids 36-91 are homologous to human G/T mismatch-specific thymine DNA glycosylase (Genbank accession number AAC50540.1), which has 5-methylcytosine glycosylase activity (Zhu *et al.*, *Nuc. Acids Res.* 28:4157-4165 (2000)). DMT polypeptides also contain a leucine zipper sequence (e.g., positions 1330-1351 of SEQ ID NO:2), that can be involved in protein-protein interactions as well as DNA binding. In addition, the amino portion of the DMT polypeptide (amino acids 43-78) is generally basic, similar to histone H1. Thus, without intending to limit the scope of the invention, it is believed this basic portion of DMT facilitates interactions with DNA and/or chromatic proteins.

In addition, amino acids 1-800 is related to the beta subunit of bacterial DNA-dependent RNA polymerases. Without intending to limit the scope of the invention, it is believed the RNA polymerase-like domain facilitates interaction of DMT with DNA.

Amino acids 1167-1368 is related to proteins in the HhH-GPD superfamily. Amino acids 1,271 to 1,304 correspond to the conserved HhH-GPD motif. The corresponding DMT sequence is DKAKDYLLSIRGLGLKSVECVRLLTLHNLAFPVD. Secondary structure prediction

(Jpred program) indicates that DMT has two alpha-helices (1,271 - 1,279 and 1,286 to 1,295) that correspond to the conserved alphaK and alphaL helices in the HhH-GPD motif of the crystallized hOGG1 DNA repair protein (Bruner et al *Nature* 403:859-866 (2000)). In between the two helices (1280 to 1285), is a hairpin with conserved glycines (G1282 and G1284). Amino acids 1286 to 1295 are related to the alphaL helix of hOGG1, which contacts the DNA backbone (Bruner *et al Nature* 403:859-866 (2000). Thus, without intending to limit the scope of the invention, it is believed this region of DMT contacts the DNA. The catalytic lysine (K1286) and aspartic acid (D1304) residues are conserved in the HhH-GPD motif of DMT. Without intending to limit the scope of the invention, by analogy to hOGG1, K1286 is predicted to displace the modified base and to promote conjugate elimination of the 3'-phosphodiester bond. Without intending to limit the scope of the invention, by analogy to hOGG1, D1304 is believed to assist the reaction by transferring protons to and from K1286.

DMT nucleic acids are a new class of plant regulatory genes that encode polypeptides with sequence identity to members of the endonuclease III genes found in a diverse collection of organisms. Endonuclease III is implicated in various DNA repair reactions. Thus proteins related to endonuclease III are likely to have a chromosomal function. DMT (SEQ ID NO:1) is most related to endonuclease III from *Deinococcus radiodurans* Genbank Accession No. AE002073 (*see*, e.g., White, O. *et al. Science* 286:1571-1577 (1999)). DMT polypeptides have glycosylase activity (i.e., the capability to cleave the base portion of a nucleotide from the sugar portion). More particularly, DMT polypeptides have demethylase activity, and in more preferred embodiments, have 5-methylcytosine glycosylase activity. Demethylation activity can be assayed *in vivo* by expressing a candidate polypeptide in the nucleus of a cell and then assaying for a change in methylation of the cell's DNA. *See, e.g.,* Vong, *et al.*, *Science* 260:1926-1928 (1993). Changes in chromosomal methylation can be measured by comparing the ability of methylation sensitive and insensitive endonucleases to cleave DNA from a cell expressing a polypeptide suspected of having demethylase or methylase activity. Alternatively, bisulfate sequencing can be used to identify which base pairs are methylated in a DNA sequence. For a discussion of both methods, see Soppe *et al.*, *Molec. Cell.* 6:791-802 (2000). *In vitro* assays to measure demethylase activity using labeled substrates are also known to those of skill in the art. *See, e.g.*, Vhu *et al.*, *Proc. Natl. Acad. Sci. USA* 97:5135-5139 (2000).

In the case of both expression of transgenes and inhibition of endogenous genes (e.g., by antisense, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only "substantially identical" to a sequence of the gene from which it was derived. As explained below, these substantially identical variants are specifically covered by the term *DMT* nucleic acid.

In the case where the inserted polynucleotide sequence is transcribed and translated to produce a functional polypeptide, one of skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same polypeptide. These variants are specifically covered by the terms "*DMT* nucleic acid". In addition, the term specifically includes those sequences substantially identical (determined as described below) with a DMT polynucleotide sequence disclosed here and that encode polypeptides that are either mutants of wild type DMT polypeptides or retain the function of the DMT polypeptide (*e.g.*, resulting from conservative substitutions of amino acids in the DMT polypeptide). In addition, variants can be those that encode dominant negative mutants as described below.

Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a

score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, e.g., the algorithm of Meyers & Miller, *Computer Applic. Biol. Sci.* 4:11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to a sequence or subsequence that has at least 40% sequence identity with a reference sequence. Alternatively, percent identity can be any integer from 40% to 100%. More preferred embodiments include at least: 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% compared to a reference sequence using the programs described herein; preferably BLAST using standard parameters, as described below. This definition also refers to the complement of a test sequence, when the test sequence has substantial identity to a reference sequence.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection.

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps.

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al., J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al, supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the

alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid.

Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

(*see, e.g.*, Creighton, *Proteins* (1984)).

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, highly stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. Low stringency conditions are generally selected to be about 15-30°C below the T_m. The

T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 55°C, sometimes 60°C, and sometimes 65°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 time background hybridization.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cased, the nucleic acids typically hybridize under moderately stringent hybridization conditions.

In the present invention, genomic DNA or cDNA comprising *DMT* nucleic acids of the invention can be identified in standard Southern blots under stringent conditions using the nucleic acid sequences disclosed here. For the purposes of this disclosure, suitable stringent conditions for such hybridizations are those which include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and at least one wash in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C to about 60°C and sometimes 65°C, for 20 minutes, or equivalent conditions. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

A further indication that two polynucleotides are substantially identical is if the reference sequence, amplified by a pair of oligonucleotide primers, can then be used as a probe under stringent hybridization conditions to isolate the test sequence from a cDNA or genomic library, or to identify the test sequence in, e.g., a northern or Southern blot.

DETAILED DESCRIPTION

This invention provides molecular strategies for controlling plant development, including methylation of chromosomal DNA, endosperm development and flowering time.

Reproduction in flowering plants involves two fertilization events in the haploid female gametophyte. One sperm nucleus fertilizes the egg to form the embryo. A second sperm nucleus fertilizes the central cell to form the endosperm, a unique tissue that supports the growth of the embryo. Fertilization also activates maternal tissue differentiation, the ovule integuments form the seed coat and the ovary forms the fruit.

The present invention is based, at least in part, on the discovery of a set of female-gametophytic mutations and the subsequent cloning of the gene involved, termed *DEMETER* (*DMT*), formally known as *ATROPOS* (*ATR*). Two mutant alleles of *DMT* disclosed here were created using a T-DNA tag, thereby disrupting an exon of the gene. The *dmt* mutations affect endosperm production, allowing for increased endosperm development. Generally, the mutant *dmt* alleles are not transmitted by the female gametophyte. Inheritance of a mutant *dmt* allele by the female gametophyte usually results in embryo abortion and endosperm overproduction, even when the pollen bears the wild-type *DMT* allele.

In contrast, transmission of *dmt* mutant alleles through the male gametophyte (*i.e.*, pollen) is ecotype-dependent in *Arabidopsis*. For instance, in some ecotypes (e.g., Columbia), transmission of *dmt* mutant alleles is less than 50%. However, in Landsberg *erecta*, transmission is almost normal.

DMT is a repressor of endosperm both before and after fertilization. DMT is both necessary and sufficient for MEDEA transcription. DMT is related to 5-methylcytosine glycosylases. DMT regulates transcription of specific target genes (i.e., MEA) by a demethylation mechanism. DMT is also required for maintaining the proper global pattern of methylation of chromosomal DNA in cells.

The isolated sequences prepared as described herein, can be used in a number of techniques, for example, to suppress or enhance endogenous DMT gene expression. Modulation of *DMT* gene expression or DMT activity in plants is particularly useful, for example, in producing embryo-less or embryo-reduced seed, seed with increased endosperm, as part of a system to generate seed, to modulate time to flowering, organ identity, size and/or number,meristem size or activity in plants, or to modulate methylation, and thus gene expression in plants. Another use is the expression of DMT

polynucleotides in animal cells, for instance as a DNA repair enzyme useful in preventing the unnatural proliferation of cells (including cancer) due to chromosomal lesions. *See, e.g.*, Bruner, *et al.*, *Nature* 403:859 (2000).

As described in more detail below, reduction of expression of DMT in plants results in a number of diverse phenotypes. Without intending to limit the invention to particular embodiments, it is belived that some of the phenotypes that are generated in plants are epigenetic mutations, i.e., effects due to differences in the methylation state of the chromosome that result in altered gene expression. Thus, DMT provides a powerful tool to develop any number of plant lines with a variety of desired phenotypes.

Isolation of DMT nucleic acids

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989).

The isolation of *DMT* nucleic acids may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed here can be used to identify the desired gene in a cDNA or genomic DNA library. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a cDNA library, mRNA is isolated from the desired organ, such as ovules, and a cDNA library which contains the *DMT* gene transcript is prepared from the mRNA. Alternatively, cDNA may be prepared from mRNA extracted from other tissues in which *DMT* genes or homologs are expressed.

The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned *DMT* gene disclosed here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Alternatively, antibodies raised against a DMT polypeptide can be used to screen an mRNA expression library.

Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of the *DMT* genes directly from genomic DNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes. For a general overview of PCR see *PCR Protocols: A Guide to Methods and Applications*. (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), *Academic Press*, San Diego (1990).

Appropriate primers and probes for identifying *DMT* sequences from plant tissues are generated from comparisons of the sequences provided here with other related genes. For instance, *DMT* can be compared to the other endonuclease III genes, such as Genbank Accession No. AE002073. Using these techniques, one of skill can identify conserved regions in the nucleic acids disclosed here to prepare the appropriate primer and probe sequences. Primers that specifically hybridize to conserved regions in *DMT* genes can be used to amplify sequences from widely divergent plant species. Appropriate primers for amplification of the genomic region or cDNA of *DMT* include the following primers:

Xba-SKEN-7; CCTCTAGAGGAATTGTCGGCAAAATCGAG
SKB-8; GGAGAGACGGTTATTGTCAACC
SKB-7; AAAAGTCTACAAGGGAGAGAGAGT
SKB-5; GTAGATGTACATACGTACC
SKEN-8; GCATCCTCCAACAAGTAACAATCCACTC
SKB-6; CACTGAGATTAATTCTTCAGACTCG
SKEN-3.5; CTCAGGCGAGTCAATGCCGGAGAACAC
SKEN-3; CGAGGGCTGATCCGGGGGATAGATATTTT
SKEN-2; CCCCCGGATCAGCCCTCGAATTC
SKEN-1; CCCCTGTCTACAAATTCACCACCTGG
SKEL-4; CTGACCCAACTGCTTCTCTTC
skes1.5; TCACCTGTTCTGAACAGACTGG
SKES-1.4; CAGCAGACGAGTCCATAATGCTCTGC
SKES-2.4; GGTTTGCCTTCCACGACCACC
SKES-1; GGAAGCCACGCAAAGCTGCAACTCAGG
SKES-2.45; GAGTTGCAGCTTTGCGTGGCTTCC

SKES2.5; TTCAGACTCAGAGTCACCTTGC

SKES-2; ACCAGCAGCCTTGCTTGGCC

SKES-3; CATGCCAGAGAAGCAGGGCTCC

SKES3.5; CGATGATACTGTCTCTTCGAGC

SKES-6; CCTCCGCCTGCTCATGCCTCAG

SKEN-4; GTCCATCAGGAGAACTTCTGTGTCAGGAT

SKES-4; GGGAACAAGTGCACCATCTCC

SKEN-6; GCTCTCATAGGGAACAAGTGCACCATCTC

SKES-5; CGCTCGCATGCACCTGGTAC

SKB-1; GGAGGGAATCGAGCAGCTAGAG

SKB-2; GAGCAGCTAAGGGACTGTTCAAACTC

SKB-3; CCAGGAATGGGATTGTCCGG

3' RACE-2; CTTGGACGGCGCTTGAGGAACC

3' RACE-1; GCCTACAAGCCAGTGGGATAG

cDNA-1; GCCAAGGACTATCTCTTGAGC

SKB-4; GGATGGACTCGAGCACTGGG

SKE2.2-4; AGAGGAGAGTGCAGACACTTTG

cDNA-3; GAGGACCCTGACGAGATCCCAAC

cDNA-9; CCATGTGTTCCCGTAGAGTCATTCC

2.2+SKE-1; ATGGAGCTCCAAGAAGGTGACATG

cDNA-5; CAGAAGTGTGGAGGGAAAGCGTCTGGC

cDNA-4; CCCTCAGACTGTTACACTCAGAAC

cDNA-2; CCCGTTGAGCGGAAAACTTCCTCTCATGGC

cDNA-7; GGAAAGGATTCGTATGTGTCCGTGG

SKEN-5; GCAATGCGTTTGCTTTCTTCCAGTCATCT

cDNA-6; GAGGAGAGCAGAGAAGCAATGCGTTTGC

cDNA-8; GTTAGAGAGAAAATAAATAACCC

2.2+SKE-3; CCCTAAACAACACCGGATACAC

The amplification conditions are typically as follows. Reaction components: 10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.001% gelatin, 200 μM dATP, 200 μM dCTP, 200 μM dGTP, 200 μM dTTP, 0.4 μM primers, and 100 units per ml Taq polymerase. Program: 96 C for 3 min., 30 cycles of 96 C for 45 sec., 50 C for 60 sec., 72 for 60 sec, followed by 72 C for 5 min.

Standard nucleic acid hybridization techniques using the conditions disclosed above can then be used to identify full-length cDNA or genomic clones.

Alternatively, a number of methods for designing modifications of polynucleotide sequences are known to those of skill in the art. For example, oligonucleotide directed mutagenesis can be used to introduce site-specific mutations in a nucleic acid sequence of interest. Examples of such techniques are found in the references above and, e.g., in Reidhaar-Olson *et al. Science*, 241:53-57 (1988) and Ausubel *et al.* Similarly, gene shuffling (Stemmer *Proc. Natl. Acad. Sci. USA* 91:10747-10751(1994); Ostermeier *et al. Proc. Natl. Acad. Sci. USA*, 96: 3562-67(1999))) can be used to introduce variation into one or more *DMT* sequences or subsequences. For example, orthologous (between species) or homologous (within a species) *DMT* nucleic acids can be interchanged, combined or shuffled to produce novel variations within the scope of the invention.

Additionally, error prone PCR can also be used to introduce variation into a nucleic acid sequence. *See*, Leung *et al.* (1989) *Technique* 1:11-15 and Caldwell *et al.* (1992) *PCR Methods Applic.* 2:28-33.

Control of DMT activity or gene expression

Since *DMT* genes are involved in controlling seed, in particular endosperm, development, inhibition of endogenous DMT activity or gene expression is useful in a number of contexts. For instance, reduction of DMT activity can be used for production of seed with enhanced endosperm. By reducing and/or eliminating DMT activity, plants with seed containing increased endosperm can be produced.

Alternatively, substantial inhibition of DMT activity can be used for production of fruit with small and/or degraded seed (referred to here as "seedless fruit") after fertilization. In many plants, particularly dicots, the endosperm is not persistent and eventually is degraded. Thus, in plants of the invention in which DMT activity is inhibited, embryo-less seed do not persist and seedless fruit are produced. For production of dicots with enhanced endosperm, the most beneficial effect may be to reduce, but not eliminate DMT activity. On the other hand, in monocots, which have persistent endosperm, it is advantageous to eliminate DMT activity.

Alternatively, plants of the invention can be used to prevent pre-harvest sprouting in seeds, especially those derived from cereals. In these plants, the endosperm persists and is the major component of the mature seed. Premature growth of embryos in stored grain causes release of degradative enzymes which digest starch and other

components of the endosperm. Plants of the present invention are useful in addressing this problem because the seeds lack an embryo and thus will not germinate.

Moreover, as discussed herein, time to flowering and DNA methylation can also be modulated by modulating DMT activity in a cell. For example, DMT can be used to modulate the amount of methylated DNA in a cell. Indeed, since expression of many genes is dependent on their methylation state, modulation of DMT activity modulates gene expression in a cell. Examples of genes whose expression is modulated by DMT include *MEDEA*.

One of skill will recognize that a number of methods can be used to modulate DMT activity or gene expression. DMT activity can be modulated in the plant cell at the gene, transcriptional, posttranscriptional, translational, or posttranslational, levels. Techniques for modulating DMT activity at each of these levels are generally well known to one of skill and are discussed briefly below.

Methods for introducing genetic mutations into plant genes are well known. For instance, seeds or other plant material can be treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are not limited to, the following: diethyl sulfate, ethylene imine, ethyl methanesulfonate and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as, for example, X-rays or gamma rays can be used.

Alternatively, homologous recombination can be used to induce targeted gene disruptions by specifically deleting or altering the *DMT* gene *in vivo* (*see, generally*, Grewal and Klar, *Genetics* 146: 1221-1238 (1997) and Xu *et al.*, *Genes Dev.* 10:2411-2422 (1996)). Homologous recombination has been demonstrated in plants (Puchta *et al.*, *Experientia* 50:277-284 (1994), Swoboda *et al.*, *EMBO J.* 13:484-489 (1994); Offringa *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 7346-7350 (1993); and Kempin *et al. Nature* 389:802-803 (1997)).

In applying homologous recombination technology to the genes of the invention, mutations in selected portions of a DMT gene sequences (including 5' upstream, 3' downstream, and intragenic regions) such as those disclosed here are made *in vitro* and then introduced into the desired plant using standard techniques. Since the efficiency of homologous recombination is known to be dependent on the vectors used, use of dicistronic gene targeting vectors as described by Mountford *et al. Proc. Natl. Acad. Sci. USA* 91:4303-4307 (1994); and Vaulont *et al. Transgenic Res.* 4:247-255 (1995) are conveniently used to increase the efficiency of selecting for altered *DMT* gene

expression in transgenic plants. The mutated gene will interact with the target wild-type gene in such a way that homologous recombination and targeted replacement of the wild-type gene will occur in transgenic plant cells, resulting in suppression of DMT activity.

Alternatively, oligonucleotides composed of a contiguous stretch of RNA and DNA residues in a duplex conformation with double hairpin caps on the ends can be used. The RNA/DNA sequence is designed to align with the sequence of the target *DMT* gene and to contain the desired nucleotide change. Introduction of the chimeric oligonucleotide on an extrachromosomal T-DNA plasmid results in efficient and specific DMT gene conversion directed by chimeric molecules in a small number of transformed plant cells. This method is described in Cole-Strauss *et al. Science* 273:1386-1389 (1996) and Yoon *et al. Proc. Natl. Acad. Sci. USA* 93:2071-2076 (1996).

Gene expression can be inactivated using recombinant DNA techniques by transforming plant cells with constructs comprising transposons or T-DNA sequences. *DMT* mutants prepared by these methods are identified according to standard techniques. For instance, mutants can be detected by PCR or by detecting the presence or absence of *DMT* mRNA, *e.g.*, by Northern blots. Mutants can also be selected by assaying for development of endosperm in the absence of fertilization.

The isolated nucleic acid sequences prepared as described herein, can also be used in a number of techniques to control endogenous *DMT* gene expression at various levels. Subsequences from the sequences disclosed here can be used to control, transcription, RNA accumulation, translation, and the like.

A number of methods can be used to inhibit gene expression in plants. For instance, antisense technology can be conveniently used. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The construct is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been suggested that antisense suppression can act at all levels of gene regulation including suppression of RNA translation (*see*, Bourque *Plant Sci. (Limerick)* 105:125-149 (1995); Pantopoulos In Progress in Nucleic Acid Research and Molecular Biology, Vol. 48. Cohn, W. E. and K. Moldave (Ed.). Academic Press, Inc.: San Diego, California, USA; London, England, UK. p. 181-238; Heiser *et al. Plant Sci. (Shannon)* 127:61-69 (1997)) and by preventing the accumulation of mRNA which encodes the protein of interest, *(see,* Baulcombe *Plant Mol. Bio.* 32:79-88 (1996); Prins and Goldbach *Arch. Virol.* 141:2259-2276 (1996);

Metzlaff *et al. Cell* 88:845-854 (1997), Sheehy et al., *Proc. Nat. Acad. Sci. USA*, 85:8805-8809 (1988), and Hiatt et al., U.S. Patent No. 4,801,340).

The nucleic acid segment to be introduced generally will be substantially identical to at least a portion of the endogenous *DMT* gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. The vectors of the present invention can be designed such that the inhibitory effect applies to other genes within a family of genes exhibiting homology or substantial homology to the target gene.

For antisense suppression, the introduced sequence also need not be full length relative to either the primary transcription product or fully processed mRNA. Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and about full length nucleotides should be used, though a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of about 500 to about 7000 nucleotides is especially preferred.

A number of gene regions can be targeted to suppress *DMT* gene expression. The targets can include, for instance, the coding regions, introns, sequences from exon/intron junctions, 5' or 3' untranslated regions, and the like. In some embodiments, the constructs can be designed to eliminate the ability of regulatory proteins to bind to *DMT* gene sequences that are required for its cell- and/or tissue-specific expression. Such transcriptional regulatory sequences can be located either 5'-, 3'-, or within the coding region of the gene and can be either promote (positive regulatory element) or repress (negative regulatory element) gene transcription. These sequences can be identified using standard deletion analysis, well known to those of skill in the art. Once the sequences are identified, an antisense construct targeting these sequences is introduced into plants to control gene transcription in particular tissue, for instance, in developing ovules and/or seed. In one embodiment, transgenic plants are selected for DMT activity that is reduced but not eliminated.

Oligonucleotide-based triple-helix formation can be used to disrupt DMT gene expression. Triplex DNA can inhibit DNA transcription and replication, generate site-specific mutations, cleave DNA, and induce homologous recombination (*see, e.g.*, Havre and Glazer *J. Virology* 67:7324-7331 (1993); Scanlon *et al. FASEB J.* 9:1288-1296 (1995); Giovannangeli *et al. Biochemistry* 35:10539-10548 (1996); Chan and Glazer *J.*

Mol. Medicine (Berlin) 75:267-282 (1997)). Triple helix DNAs can be used to target the same sequences identified for antisense regulation.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of *DMT* genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs. Thus, ribozymes can be used to target the same sequences identified for antisense regulation.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Zhao and Pick *Nature* 365:448-451 (1993); Eastham and Ahlering *J. Urology* 156:1186-1188 (1996); Sokol and Murray *Transgenic Res.* 5:363-371 (1996); Sun *et al. Mol. Biotechnology* 7:241-251 (1997); and Haseloff *et al. Nature*, 334:585-591 (1988).

Another method of suppression is sense cosuppression. Introduction of nucleic acid configured in the sense orientation has been recently shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes (*see,* Assaad *et al. Plant Mol. Bio.* 22:1067-1085 (1993); Flavell *Proc. Natl. Acad. Sci. USA* 91:3490-3496 (1994); Stam *et al. Annals Bot.* 79:3-12 (1997); Napoli et al., *The Plant Cell* 2:279-289 (1990); and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184).

The suppressive effect may occur where the introduced sequence contains no coding sequence *per se*, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed. This minimal identity will typically be greater than about 65%, but a higher identity might exert a more effective repression of expression of the endogenous

sequences. Substantially greater identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect should apply to any other proteins within a similar family of genes exhibiting homology or substantial homology.

For sense suppression, the introduced sequence, needing less than absolute identity, also need not be full length, relative to either the primary transcription product or fully processed mRNA. This may be preferred to avoid concurrent production of some plants that are overexpressers. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Normally, a sequence of the size ranges noted above for antisense regulation is used. In addition, the same gene regions noted for antisense regulation can be targeted using cosuppression technologies.

In a preferred embodiment, expression of a nucleic acid of interest can be suppressed by the simultaneous expression of both sense and antisense constructs (Waterhouse *et al.*, *Proc. Natl. Acad. Sci. USA* 95:13959-13964 (1998). *See also* Tabara *et al. Science* 282:430-431 (1998).

Alternatively, *DMT* activity may be modulated by eliminating the proteins that are required for *DMT* cell-specific gene expression. Thus, expression of regulatory proteins and/or the sequences that control *DMT* gene expression can be modulated using the methods described here.

Another method is use of engineered tRNA suppression of *DMT* mRNA translation. This method involves the use of suppressor tRNAs to transactivate target genes containing premature stop codons (*see*, Betzner *et al. Plant J.*11:587-595 (1997); and Choisne *et al. Plant J.*11:597-604 (1997). A plant line containing a constitutively expressed *DMT* gene that contains an amber stop codon is first created. Multiple lines of plants, each containing tRNA suppressor gene constructs under the direction of cell-type specific promoters are also generated. The tRNA gene construct is then crossed into the *DMT* line to activate *DMT* activity in a targeted manner. These tRNA suppressor lines could also be used to target the expression of any type of gene to the same cell or tissue types.

DMT proteins may form homogeneous or heterologous complexes *in vivo*. Thus, production of dominant-negative forms of *DMT* polypeptides that are defective in their abilities to bind to other proteins in the complex is a convenient means to inhibit

endogenous DMT activity. This approach involves transformation of plants with constructs encoding mutant DMT polypeptides that form defective complexes and thereby prevent the complex from forming properly. The mutant polypeptide may vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain. Use of dominant negative mutants to inactivate target genes is described in Mizukami *et al. Plant Cell* 8:831-845 (1996).

Another strategy to affect the ability of a DMT protein to interact with itself or with other proteins involves the use of antibodies specific to DMT. In this method cell-specific expression of DMT-specific Abs is used inactivate functional domains through antibody:antigen recognition (*see*, Hupp *et al. Cell* 83:237-245 (1995)).

After plants with reduced DMT activity are identified, a recombinant construct capable of expressing low levels of DMT in embryos can be introduced using the methods discussed below. In this fashion, the level of DMT activity can be regulated to produce preferred plant phenotypes. For example, a relatively weak promoter such as the ubiquitin promoter (see, e.g., Garbarino *et al. Plant Physiol.* 109(4):1371-8 (1995); Christensen *et al Transgenic Res.* 5(3):213-8 (1996); and Holtorf *et al. Plant. Mol. Biol.* 29(4):637-46 (1995)) is useful to produce plants with reduced levels of DMT activity or expression. Such plants are useful for producing, for instance, plants that produce seed with enhanced endosperm.

Use of nucleic acids of the invention to enhance DMT gene expression

Isolated sequences prepared as described herein can also be introduced into a plant cell, thereby modulating expression of a particular *DMT* nucleic acid to enhance or increase endogenous gene expression. For instance, without being bound to any theory, in light of DMT's relation to Exonuclease III and DNA glycosylases, applicants believe that DMT binds DNA or chromatin and acts to modulate transcription by modulating the methylation state of DNA. Enhanced expression can therefore be used to control plant morphology by controlling expression of genes under DMT's control, such as *MEDEA*, in desired tissues or cells. Enhanced expression can also be used, for instance, to increase vegetative growth by preventing the plant from setting seed. Where overexpression of a gene is desired, the desired gene from a different species may be used to decrease potential sense suppression effects.

Moreover, as discussed herein, time to flowering and DNA methylation can also be modulated by modulating DMT activity in a cell. For example, increased expression of DMT in a plant results in delayed time to flowering. Similarly, DMT can be used to modulate the amount of methylated DNA in a cell. Indeed, since expression of many genes is dependent on their methylation state, modulation of DMT activity modulates gene expression in a cell. Examples of genes whose expression is modulated by DMT include *MEDEA*.

One of skill will recognize that the polypeptides encoded by the genes of the invention, like other proteins, have different domains that perform different functions. Thus, the gene sequences need not be full length, so long as the desired functional domain of the protein is expressed.

Modified protein chains can also be readily designed utilizing various recombinant DNA techniques well known to those skilled in the art and described in detail, below. For example, the chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

Preparation of recombinant vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of flowering plant species are well known and described in the technical and scientific literature. See, for example, Weising *et al. Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for overexpression, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of

Agrobacterium tumafaciens, and other transcription initiation regions from various plant genes known to those of skill. Such genes include for example, *ACT11* from *Arabidopsis* (Huang *et al. Plant Mol. Biol.* 33:125-139 (1996)), *Cat3* from *Arabidopsis* (GenBank No. U43147, Zhong *et al.*, *Mol. Gen. Genet.* 251:196-203 (1996)), the gene encoding stearoyl-acyl carrier protein desaturase from *Brassica napus* (Genbank No. X74782, Solocombe *et al. Plant Physiol.* 104:1167-1176 (1994)), *GPc1* from maize (GenBank No. X15596, Martinez *et al. J. Mol. Biol* 208:551-565 (1989)), and *Gpc2* from maize (GenBank No. U45855, Manjunath *et al.*, *Plant Mol. Biol.* 33:97-112 (1997)).

Alternatively, the plant promoter may direct expression of the *DMT* nucleic acid in a specific tissue or may be otherwise under more precise environmental or developmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light. Such promoters are referred to here as "inducible" or "tissue-specific" promoters. One of skill will recognize that a tissue-specific promoter may drive expression of operably linked sequences in tissues other than the target tissue. Thus, as used herein a tissue-specific promoter is one that drives expression preferentially in the target tissue, but may also lead to some expression in other tissues as well.

Examples of promoters under developmental control include promoters that initiate transcription only (or primarily only) in certain tissues, such as fruit, seeds, or flowers. Promoters that direct expression of nucleic acids in ovules, flowers or seeds are particularly useful in the present invention. As used herein a seed-specific promoter is one which directs expression in seed tissues, such promoters may be, for example, ovule-specific (which includes promoters which direct expression in maternal tissues or the female gametophyte, such as egg cells or the central cell), embryo-specific, endosperm-specific, integument-specific, seed coat-specific, or some combination thereof. Examples include a promoter from the ovule-specific *BEL1* gene described in Reiser *et al. Cell* 83:735-742 (1995) (GenBank No. U39944). Other suitable seed specific promoters are derived from the following genes: *MAC1* from maize (Sheridan *et al. Genetics* 142:1009-1020 (1996), *Cat3* from maize (GenBank No. L05934, Abler *et al. Plant Mol. Biol.* 22:10131-1038 (1993), the gene encoding oleosin 18kD from maize (GenBank No. J05212, Lee *et al. Plant Mol. Biol.* 26:1981-1987 (1994)), vivparous-1 from *Arabidopsis* (Genbank No. U93215), the gene encoding oleosin from *Arabidopsis* (Genbank No. Z17657), Atmyc1 from *Arabidopsis* (Urao *et al.Plant Mol. Biol.* 32:571-576 (1996), the 2s seed storage protein gene family from *Arabidopsis* (Conceicao *et al. Plant* 5:493-505

(1994)) the gene encoding oleosin 20kD from *Brassica napus* (GenBank No. M63985), *napA* from *Brassica napus* (GenBank No. J02798, Josefsson *et al. JBL* 26:12196-1301 (1987), the napin gene family from *Brassica napus* (Sjodahl *et al. Planta* 197:264-271 (1995), the gene encoding the 2S storage protein from *Brassica napus* (Dasgupta *et al. Gene* 133:301-302 (1993)), the genes encoding oleosin A (Genbank No. U09118) and oleosin B (Genbank No. U09119) from soybean and the gene encoding low molecular weight sulphur rich protein from soybean (Choi *et al. Mol Gen, Genet.* 246:266-268 (1995)).

In addition, the promoter sequences from the *DMT* genes disclosed here can be used to drive expression of the *DMT* polynucleotides of the invention or heterologous sequences. The sequences of the promoters are identified below.

If proper polypeptide expression is desired, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences (*e.g.*, promoters or coding regions) from genes of the invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

Promoter and Enhancer Nucleic Acids of the Invention

The present invention provides polynucleotides useful as promoters and enhancers. The invention also provides methods of targeting heterologous polypeptides to a female gametophyte of a plant, including, e.g., the polar nuclei, the eggs and synergids and central cells. Promoter polynucleotides of the invention include, for example, sequences and subsequences of the DMT 5' flanking DNA (SEQ ID NO:3), the 5' UTR region (SEQ ID NO:6) and the 3' flanking region (SEQ ID NO:4). In some embodiments, the promoter sequences are operably linked to the 5' end of the DMT coding region, which is in turn fused to a polynucleotide of interest, typically encoding a polypeptide. An exemplary promoter sequence includes the last 3424 nucleotides of SEQ ID NO:3 linked to the first 1478 nucleotides of SEQ ID NO:5. In some embodiments, a further 444 nucleotides (e.g., the first 444 nucleotides of the DMT coding region) are incorporated into the promoter. In some embodiments, the promoter

sequences of the invention specifically direct expression of polynucleotides to the female gametophyte and does not direct expression in tissues following fertilization.

Production of transgenic plants

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al. *Embo J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm et al. *Proc. Natl. Acad. Sci. USA* 82:5824 (1985). Ballistic transformation techniques are described in Klein et al. *Nature* 327:70-73 (1987).

Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria. *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. *See*, for example Horsch et al. *Science* 233:496-498 (1984), and Fraley et al. *Proc. Natl. Acad. Sci. USA* 80:4803 (1983).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype such as increased seed mass. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture, Handbook of Plant Cell Culture*, pp. 124-176, MacMillilan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants,

organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* 38:467-486 (1987).

The nucleic acids of the invention can be used to confer desired traits on essentially any plant. Thus, the invention has use over a broad range of plants, including species from the genera *Anacardium, Arachis, Asparagus, Atropa, Avena, Brassica, Citrus, Citrullus, Capsicum, Carthamus, Cocos, Coffea, Cucumis, Cucurbita, Daucus, Elaeis, Fragaria, Glycine, Gossypium, Helianthus, Heterocallis, Hordeum, Hyoscyamus, Lactuca, Linum, Lolium, Lupinus, Lycopersicon, Malus, Manihot, Majorana, Medicago, Nicotiana, Olea, Oryza, Panieum, Pannesetum, Persea, Phaseolus, Pistachia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Secale, Senecio, Sinapis, Solanum, Sorghum, Theobromus, Trigonella, Triticum, Vicia, Vitis, Vigna,* and *Zea.*

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

Seed obtained from plants of the present invention can be analyzed according to well known procedures to identify plants with the desired trait. If antisense or other techniques are used to control DMT gene expression, Northern blot analysis can be used to screen for desired plants. In addition, the presence of fertilization independent reproductive development can be detected. Plants can be screened, for instance, for the ability to form embryo-less seed, form seed that abort after fertilization, or set fruit in the absence of fertilization. These procedures will depend, part on the particular plant species being used, but will be carried out according to methods well known to those of skill.

DMT Mutations, Fragments And Fusions

As discussed above, DMT polynucleotides and polypeptides are not limited to the sequences disclosed herein. Those of skill in the art that conservative amino acid substitutions, as well as amino acid additions or deletions may not result in any change in biological activity. Moreover, sequence variants with at least one modulated biological activity of DMT are also contemplated. For example, at least one DMT activity can be increased or decreased by introduction of single or multiple amino acid changes from the sequences disclosed herein. Those of skill in the art will recognize that conservative amino acid substitutions in important functional domains are typically

useful in generating more active DMT polypeptides. Conversely, non-conservative substitutions of amino acid residues in functional domains, such as the HhH region of DMT (e.g., amino acids 1271-1304 of SEQ ID NO:2) are likely to disrupt at least one biological activity such as DNA binding. In some embodiments, the fragments of the invention consist of about 100, 200, 300 400, 500, 600, 700, 800, 900, or 1000 amino acids.

Alternatively, fragments of the sequences disclosed herein are contemplated. In some preferred embodiments, the polypeptide fragments have at least one biological activity of DMT. For example, amino acid sequences comprising DMT domain B represent polypeptide fragments with glycosylase or demethylase activity. In some embodiments, a fragment comprising amino acids 1167-1404, 1192-1404, 1192-1368 or 1167-1368 of SEQ ID NO:2 have glycosylase activity.

Mutations, fragments and fusions are also useful as dominant negative mutations. For instance, different regions of the DMT protein are responsible for different biological activities. Thus, mutation or deletion of one functional domain can eliminate one but not all activities. For example, mutation or deletion of the DNA binding domain may result in proteins that interact with proteins necessary for DMT function, effectively titrating out those proteins and preventing an active DMT protein from acting. Similarly, DMT fragments comprising the DNA binding portion of the protein with an inactive enzymatic domain or lacking an enzymatic domain are also useful as dominant negative mutants by competing with active DMT polypeptides for DNA binding sites. As described herein, domains of DMT that can be modulated include: the leucine zipper, nuclear localization sequence, HhH domain, the aspartic acid of the GPD domain, as well a DMT domains A, B or C. Without intending to limit the scope of the invention, based on the data provided herein, DMT has glycosylase and demethylase activity and is a DNA repair enzyme.

Targeting the polypeptides of the invention to chromosomal regions

Without intending to limit the scope of the invention, based on the data provided herein, it is believed that DMT has glycosylase and/or demethylase activity and is a DNA repair enzyme. DNA methylation plays an important role in the repression of gene transcription during animal development including embryogenesis, myogenesis and blood cell development. Methylated DNA is recognized by MeCP2 which inturn represses gene transcription by recruiting the Sin3 repressor complex that contains

catalytically active histone deacetylase (Jones *et al. Nature Genetics* 19(2):187-191 (1998)). Histone H3 and H4 deacetylation contributes to the formation of transcriptionally inactive chromatin. Thus, DMT can be used for the purpose of modulating the activity of target genes through chromatin architecture in animal cells as well as plant cells. For example, in some embodiments, DMT is used to catalytically remove 5-MeC from target gene DNA in several ways: e.g., (1) by fusing DMT to a sequence specific DNA binding protein, or (2) by fusing DMT to a subunit of the target repressor complex such as MeCP2 or Sin3. When combined with cell, tissue, or developmentally specific promoters DMT can be used to modulate specific sets of target genes.

In addition, reactive oxygen species, partially reduced species that are produced as intermediates of aerobic respiration, are powerful oxidizing agents that escape the mitochondria and attach vial cellular components. Ionizing radiation and other agents that generate free radicals also produce reactive oxygen species that can attack the genome and cause lesions that are thought to have a key role in in causing cancer and ageing. For example, 7,8-dihydro-8-oxoguanine (oxoG) is a very deleterious adduct generated by oxidation of the guanine base in DNA. The oxoG protein can pair with either cytosine or adenine during DNA replication. Thus, oxoG residues in DNA give rise to G/C to T/A transversion mutations. These transversions are common somatic mutations found in human cancers. HhH-GPD enzymes, such as those described herein, represent a defense against oxoG by catalysing the expulsion of the oxoG. Thus, in some embodiments, enhanced DMT activity is a method to reduce the incidence of mutations in animal cells. Also, DMT can be used to catalytically remove oxoG from a target gene by fusing DMT to a sequence specific DNA binding protein. When combined with a cell, tissue, or developmentally specific promoters DMT can be used to modulate repair of target genes.

As described above, the polypeptides of the invention can be targeted to chromosomal regions of interest by linking the polypeptides of the invention, including fragments with demethylase activity, to a DNA-binding domain that binds a target sequence. For example, it is known that an enzyme that methylates DNA (Dam methylase) can be targeted to specific sites in the genome (B.V. Steensel and S. Henikoff, *Nature Biotechnology* 18:424-428 (2000)). Specifically, the methylase was tethered to the DNA-binding domain of GAL4. When recombinant GAL4-methylase protein was expressed in transgenic Drosophila, targeted methylation occurred in a region of a few

kilobases surrounding the GAL4 DNA binding sequence. In a analogous fashion, DMT, or a portion of DMT that has biological activity (e.g., a portion containing the HhH-GpD motif amino acids such as 1167 to 1368 of SEQ ID NO:2), can be tethered (e.g., as a translational fusion or chemically linked) to proteins that interact at specific sites in the genome. As a result, specific targeted regions of the genome are hypomethylated by DMT. As discussed above, typically hypomethylation promotes transcription of genes (S. E. Jacobsen, *Current Biology* 9, 617 (1999). The invention provides compositions and methods for methylation of a desired area of the chromosome by targeting DMT to those regions. Thus, these embodiments provide additional ways to activate transcription of a desired gene in a targeted chromosomal region.

The following Examples are offered by way of illustration, not limitation.

EXAMPLE

Example 1:

This example shows the characterization of *dmt* mutant plants and the isolation of *DMT*.

Arabidopsis plants were transformed by infiltrating them with Agrobacterium containing the SKI15 T-DNA vector (generously provided by D. Weigel (Salk Institute, La Jolla, CA)). T1 seeds were harvested. The SKI15 vector has the bialaphos resistance (BAR) gene that allowed us to directly select transgenic plants in soil after spraying with the commercially available herbicide, Basta. Siliques from approximately 5,000 Basta resistant plants were opened, and those displaying approximately 50% seed abortion were identified.

Two lines, B13 and B33, were identified for further characterization. Genetic analysis of the mutants revealed that the *dmt* mutants were female sterile. Male fertility, however, depended on the genetic background of the mutant alleles. For instance, in the Columbia background, transmission of the *dmt* mutation is less than 50%. However, in the Landsberg *erecta* background, transmission through the male was almost normal.

Molecular analysis confirmed that the two mutations were allelic. For example, both the B13 and B33 alleles carry the SKI15 T-DNA within a DMT exon, confirming that disruption of the *DMT* gene resulted in the observed B13 and B33 phenotypes.

Sub a7 5'- and 3'- RACE were used to delineate the 5'- and 3'-ends of the cDNA, respectively. 5'-RACE was carried out using reagents and protocols provided by 5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0, GIBCO BRL, LIFE TECHNOLOGIES, Grand Island, NY and Marathon cDNA Amplification Kit, Clontech, Palo Alto, CA. Final gene specific 5'-RACE primers were SKES-4 (GGGAACAAGTGCACCATCTCC) and SKES3.5 (CGATGATACTGTCTCTTCGAGC). 3'-RACE was carried out using reagents and protocols provided by Marathon cDNA Amplification Kit, Clontech, Palo Alto. Final gene-specific 3' end was obtained from cDNA library screening.

The nucleotide sequence of the genomic copy of *DMT* was also determined (SEQ ID NO:1). The 5'-end of the DMT RNA is located at position 3,425 of SEQ ID NO:1. The position of the 3'-end of the DMT RNA is at position 12,504 of SEQ ID NO:1. The position of the ATG translation initiation codon is at position 4,903 of SEQ ID NO:1. The position of the TAA translation termination codon is at position 12,321 of SEQ ID NO:1.

A portion of the *DMT* polynucleotide sequence, including the first exon, is encompassed by the bacterial artificial chromosome (BAC) clone T9J15TRB. For example, sequences 3820-4299, 4319-4558, 4546-5025 and 9320-9777 of SEQ ID NO:1 were previously determined using the BAC clone as a template. Moreover, a separate independently sequenced region (Bork, C. *et al Gene* 28:147-153 (1998)) also overlaps the *DMT* sequence at positions 11,087 to 12,785 of SEQ ID NO:1.

The predicted DMT protein has 1,729 amino acids. This sequence was compared to known protein sequences using BLAST and revealed homology to several Endonuclease III proteins. The highest homology was to the Endonuclease III protein from *Deinococcus radiodurans*, Genbank Accession No. AE002073 (*see*, e.g., White, O. *et al. Science* 286:1571-1577 (1999)). Other DMT motifs include two consecutive nuclear localization signals at positions 43-60 and 61-78 and a leucine zipper at positions 1330-1351.

Example 2:

This example provides further evidence that mutant phenotypes are caused by loss-of-function mutations.

A new allele, *dmt-3*, was obtained. The *dmt-3* allele was caused by insertion of the simple pD991 T-DNA vector (M. R. Sussman, *et al.*, *Plant Physiol.*

124:1465 (2000)) into the 2nd exon of the *DMT* gene. In contrast, the previous two alleles, *dmt-1* and *dmt-2*, were caused by insertion of the activation T-DNA vector, SKI015 vector. The mutant phenotypes generated by all three *dmt* alleles are the same. Because pD991 does not have activation sequences, it suggests that all three mutant alleles are loss-of-function alleles. Consistent with this conclusion, seed abortion can be rescued with a transgene with 3,373 base pairs of 5'-*DMT* flanking sequences plus 1,478 base pairs of 5-'UTR ligated to a cDNA encoding the full-length DMT polypeptide (i.e., *DMTp::DMT*). Thus, *dmt/DMT* heterozygous plants that are hemizygous for the *DMTp::DMT* transgene displayed 25% seed abortion. Control *dmt/DMT* plants displayed 50% seed abortion.

Example 3:

This example shows that *DMT* is necessary and sufficient for *MEA* gene expression.

As discussed above, when fertilization of *dmt/dmt* homozygous mutant flowers was prevented, fertilization-independent endosperm development was observed. This is very similar to when fertilization of mutant *mea* flowers is prevented. Thus, before fertilization, both *DMT* and *MEA*, a polycomb protein (T. Kiyosue *et al.*, *Proc. Natl. Acad. Sci. USA* 96:4186 (1999)), prevent the central cell of the female gametophyte from forming an endosperm. This is consistent with *DMT* being a positive regulator of MEDEA (MEA).

As further evidence of this relationship, *MEA* RNA accumulates in immature floral (IF) buds and open flowers (OF). However, in *dmt/dmt* mutant plants there was no detectable *MEA* RNA. Thus, *DMT* is necessary for *MEA* gene expression.

In addition, we have generated plants with a transgene, *CaMV::DMT*, designed to overexpress *DMT*. The full-length *DMT* cDNA was ligated to the constitutive cauliflower mosaic virus promoter, *CaMV* (S. G. Rogers, H. J. Klee, R. B. Horsch, R. T. Fraley, *Meth Enzymol* 153:253 (1987)). In control wild type plants, the *DMT* and *MEA* genes were not significantly expressed in the leaf. However, in *35S::DMT* plants, both *DMT* and *MEA* RNA level increased significantly. This shows that *DMT* is sufficient to induce *MEA* gene expression in the leaf.

Example 4:

This example shows that DMT is a member of the HhH-GPD superfamily of DNA repair enzymes.

A BLAST search, followed by a conserved domain search, revealed that DMT is highly related to the HhH-GPD superfamily of base excision DNA repair proteins (i.e., score of 70.1, E-value of $8e^{-13}$). This family contains a diverse range of structurally related DNA repair proteins. The superfamily is called the HhH-GPD family after its hallmark helix-hairpin-helix and Gly/Pro rich loop followed by a conserved aspartate (S. D. Bruner, *et al.*, *Nature* 403:859 (2000)). This includes endonuclease III (EC:4.9.99.18), 8-oxoguanine DNA glycosylases (i.e., yeast OGG1), the thymine DNA glycosylase of methyl-CPG binding protein MBD4 (B. Hendrich, *et al.*, *Nature* 401:301 (1999)), and DNA-3-methyladenine glycosylase II (EC:3.2.2.21). The predicted amino acid sequence of DMT contains many of the conserved amino acids of this superfamily. The hallmark of the superfamily of base-excision DNA repair proteins is a helix-hairpin-helix structural element followed by a Gly/Pro-rich loop and a conserved aspartic acid (i.e., HhH-GPD motif). The DMT polypeptide is 1,729 amino acids in length. Amino acids 1,271 to 1,304 correspond to the conserved HhH-GPD motif. The DMT sequence is DKAKDYLLSIRGLGLKSVECVRLLTLHNLAFPVD. The catalytic lysine (K1286) and aspartic acid (D1304) residues are conserved in the HhH-GPD motif of DMT. Secondary structure prediction (Jpred program) indicates that DMT has two alpha-helices (amino acids 1,271 - 1,279 and 1,286 to 1,295) that correspond to the conserved alphaK and alphaL helices in the HhH-GPD motif of the crystallized hOGG1 DNA repair protein (Bruner *et al Nature* 403:859-866 (2000)).

The Arabidopsis *DMT* coding sequences were also used to identify homologous sequences in both public and proprietary databases using both the BLAST and PSI-BLAST computer algorithms. This analysis revealed amino acid sequences from several plant species, including wheat, maize, rice, soybean and Arabidopsis (SEQ ID NOs: 7-29). Based on these sequences, the following consensus sequences for DMT were determined:

DMT Domain A

KV<1>(I,l)D(D,p)(E,v)T<3>W<1>(L,v)L(M,l)(E,d)<0-2>D(K,e)<1>(K,t)<1>(K,a)(W,k)(W,l)<1>(E,k)ER<2>F<1>(G,t)R<1>(D,n)(S,l)FI(A,n)RM(H,r)<1>(V,l)QG(D,n)R<1>R<1>(P,q)WKGSVVDSV(I,v)GVFLTQN(V,t)D(H,y)(L,s)SS(S,n)A(F,y)M<1>(L,v)A(A,s)<1>FP

DMT Domain B

W(D,n)<1>(L,f)R<5>E<3-6>D(S,t)<1>(D,n)(Y,w)<3>R<10>I<2>RG(M,q)(N,f)<2>L(A,s)<1>RI<2-12>FL<3>V<2>(H,n)G<1>IDLEWLR<2>(P,d)(P,s)(D,h)<1>(A,v)K<1>(Y,f)LL(S,e)(I,f)<1>G(L,i)GLKS(V,a)ECVRLL<1>L(H,k)<2>AFPVDTNVGRI(A,c)VR(M,l)G(W,l)VPL(Q,e)PLP<2>(L,v)Q(L,m)H(L,q)L(E,f)<1>YP<1>(L,m)(E,d)(S,n)(I,v)QK(F,y)LWPRLCKL(D,p)Q<1>TLYELHY(Q,h)(L,m)ITFGK<0-2>FCTK<2>PNCNACPM(R,k)<0-2>EC(R,k)(H,y)(F,y)(A,s)SA<1>(A,v)<0-10>S(A,s)(R,k)<1>(A,l)L(P,e)<1>(P,t)

DMT Domain C.

P(I,l)(I,v)E(E,f)P<1>(S,t)P<2-5>E<0-15>(D,a)IE(D,e)<4-23>(I,v)P<1>I<1>(L,f)(N,d)<8-17>(S,a)<1>(A,d)LV<8>(I,l)P<2-5>(K,r)(L,m)K<4>LRTEH<1>V(Y,f)(E,v)LPD<1>H<1>(L,i)L(E,k)<1>(D,e)D(P,i)<2>YLL(A,s)IW(T,q)P(G,d)(E,g)<6-8>(P,s)<3>C<6-10>(M,l)C<4>C<2>C<3>(R,k)E<5>(V,f)RGT(L,i)L<0-22>(L,v)FADH<1>(S,t)(S,r)<2>PI<3>(R,t)<3>(W,k)<1>L<1>(R,k)R<4>G(T,s)(S,t)<2>(S,t)I(F,c)(R,k)(G,l)L<1>(T,v)<2>I<2>(C,n)F(W,q)<1>G(F,y)(V,l)C(V,l)R<1>F(E,d)<3>(R,g)<1>P(R,k)<1>L<2>(R,h)LH<2>(A,v)SK

The first consensus sequence listed above corresponds to amino acid positions 586 through 937 of SEQ ID NO:2. The second consensus sequence listed above corresponds to amino acid positions 1117 through 1722 of SEQ ID NO:2. The consensus sequence provides amino acid sequences by position using single letter amino acid abbreviations. Numbers in carrots ("<" or ">") refer to amino acid positions where there is no consensus and which therefore, can be any amino acid. Amino acid abbreviations in parentheses indicate alternative amino acids at the same position. Capitalized letters refer to predominant consensus amino acids and lower case letters refer to amino acids that are commonly found in DMT sequences, but are not predominant.

Example 5:

This example demonstrates the relationship between DNA repair and demethylation.

For many years, attention was focused on the ability of DNA glycosylases to repair DNA. For example, glycosylases are involved in the repair of G/T mismatched bases by depurinating the thymidine base moiety. Recently it was shown that avian (B.

Zhu *et al.*, *Proc. Natl. Acad. Sci. USA* 97:5135 (2000)) and mammalian (B. Zhu *et al.*, *Nucl. Acid Res.* 28:4157 (2000)). G/T mismatch DNA glycosylases also have 5-methylcytosine-DNA glycosylase activity. That is, these enzymes are demethylases that remove 5-methylcytosine that is later replaced by cytosine. Without intending to limit the scope of the invention, it is believed that as a member of this superfamily, *DMT* is a demethylase (i.e., 5-methylcytosine glycosylase).

The methylation (i.e., amount of 5-methylcytosine) state of a gene can have a profound effect on its expression. In general, hypomethylation is associated with elevated gene expression, whereas hypermethylation is associated with decreased gene expression (S. E. Jacobsen, *Current Biology* 9:617 (1999)). Thus, it is likely that *DMT* activates *MEA* gene expression by reducing its level of methylation.

Mutations in the *DDM1* gene in Arabidopsis reduce by 70% the overall genome cytosine methylation (E. J. Finnegan, *et al.*, *Proc. Natl. Acad. Sci. USA* 93:8449 (1996); M. J. Ronemus, *et al.*, *Science* 273:654 (1996)). Such plants develop a number of phenotypic abnormalities including floral phenotypes (T. Kakutani, *et al.*, *Proc. Natl. Acad. Sci. USA* 93:12406 (1996)). Similarly, phenotypic abnormalities have been observed developing in *dmt/dmt* homozygous plants that affect petal number, floral organ fusion, and floral organ identity. Moreover, independent *CaMV::DMT* transgenic lines that overexpress DMT frequently are late-flowering. This is particularly interesting because late flowering of *ddm1* plants was shown to be due to hypomethylation of the *FWA* gene (W. J. J. Soppe *et al.*, *Mol Cell* 6:791 (2000)). Thus, without intending to limit the scope of the invention, it is believed that both *ddm1* loss-of-function mutations and overexpression of *DMT* (i.e., *CaMV::DMT*) may result in genome hypomethylation.

Example 6:

This example demonstrates targeting gene expression to the female gametophyte using a DMT promoter sequence.

DMT RNA accumulates in many plant organs such as immature flowers, mature flowers, open flowers, stems and to a lesser extent, leaves. To understand the spatial and temporal regulation of *DMT* RNA accumulation, the expression of the DMT promoter fused to reporter genes was analyzed. We fused 2,282 base pairs of *5'-DMT* sequences, the full-length *5'-UTR* (1,478 base pairs), 444 base pairs of *DMT* coding sequences that contain a nuclear localization signal to two reporter genes, the green

fluorescent protein (*GFP*; (Y. Niwa, *et al.*, *Plant J.* 18:455 (1999))) and β-glucuronidase (*GUS*; (R. A. Jefferson, T. A. Kavanagh, M. V. Bevan, *EMBO J* 6:3901 (1987))). Reporter gene expression was observed in the developing female gametophyte, in the polar nuclei before they fuse, in the egg and synergids, and in the central cell. Expression was not detected after fertilization. Thus, this promoter is useful for targeting gene expression to the female gametophyte.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

SEQ ID NO:1 DMT genomic sequence

DMT genomic sequence (12,785 bp)

AAGCTTAAAGCTACCAACATCGAATTTAGTAAAAGACCCATGATTTGAAATTGGAATTGTCGG
CAAAATCGAGAAGATAT
AGAGCCGACACGGGAACAGTGAAAACCACAAAGCGCGTAAGAATGAAACAGTGGGAGAAGG
AAGAGAGAATCTTACCGAT
CATTCGAGGGAAAAGATGGGAATCAGAGAAAAATCTGGAAAAAAAGAAATTAAGAGAAAGA
GAGAGAAGAAAGTGAGGAG
GAAGATGCAGTGAAGACTGCTATAGCCACATCCCACATGGTGTGATGAGAGAGAGAGAGAGA
GAGGTTAAAGCAGCAAAT
TGTGGAGAGATAAAGAGAGAGAGAGACTGAGCGAGTCAAGTTCGTCGTCGTGTTTAAAAGAA
AGAATCCTATATTTGCCT
TTTTCTTTACTACTTTATTTTCAGACTATTTGCTTATTTTGCCTCAAACTTTTTTGATTGTCACTT
TTCGATCCTAAAGT
GTTTGACAATTTACCTGCCTTTTTCTCCAAGAAAAATCAGAACAGACCACAGCAAATTTATGTA
TTTTCTATTAAAAAAG
AAAGAAAGAATTCATATTACTTATAGAATTAAAAGCTAAGCAGTTGAAAACGTGAAAGCAGA
ATTTCTAAAAAAAATAGT
AAACTGCTACAAACTTATTTATGTGTATATAACATATCTATAAAGAAACTCAAATATATGATA
AATCATTTTAACAAAAT
TTCTATGAAATTATAATAAAAAAAGTCACTTTTGACACTTAAAAGGTTGACAATAACCGTCTCT
CCAAAAAAAAAATCAAA
ACATTTATAATTTCTAAAACTATGGTGTAATTTTGCTGAAATCAAAAAGAAAAGAAGGATTTC
TATATCATAAGTTTCAT
TATTGTATCAAACTTTCAAATTTCATGTAATTTGAAAGGAAAAAAATTAAGATATAATGTTGTT
TTTGTTTCTTATGTTA
CATTTTCATGGAATATATATTCATAACAAAAAATGTATTTTAATATGATGAGAGATTACCATCC
AAAAGGTCGAACTTAT
ATAAAACAAGTTAATAACTAAACAATACATGTGATCACAATCAATGACAGTTTTGATCTTAAA
ATAGAAATGATTGAGCA
AACCTCAAAAATGTCTTCTTAGGATCACAAAATCTTTCCTTTAGCTTATTAAAGCCGGGAGTTC
AACTCTCTCTCCCTTG
TAGACTTTTTGTTTTCAAATCTTTTTCTTTCAAAAAATCAATAATTAGTTAATGGGCATAATATT
TGGTTTTAATTAAGT
CCATAGATTTTTTAGGACCATCTCTAATCACGACAAATATCCTAAATTGTAACACATTTAAAAC
TTAAAAGTATTGCATT
CACAATCCTTAAAATATATATATATATATATATATATATATATATATATATATATATGAAAGTTAT
ATAGAAACGATAACTC

CTTACTCAACAATTAGCCCAAAAAAACATCCATAATGCATTTAAACTAGGAATTTTAACAAAC
TCAAATAGGTTGGTAGT
TAAAAAAAAACAAATAGTAGATGTACATACGTACCTTTAAAAATATATACTCATATCGAAAGT
TTTAAATTTTGCGAAAT
TAAATACATTTATCTATCAATTAAAATACATTTAATAATGCATAATTCTGTAATATCTATCTTT
AATTTCCATATAGAAC
CAAAACAAAATAAACATATCAAATAGTTTTAACTTAACAAAAACGTTAGGGAAAAGTTGACCT
AACTAGCTTGATTGACG
TTGAACTTGTCAATGCGAAAGCGATATTTCCAATATATACTACATGTAGTATTATTTATATGGA
AGTTTCTAAAAAGGTG
TTGAGTGGATTGTTACTTGTTGGAGGATGCTATTTTTTCCTTCTTGCCATAATATTTTACGAGTA
TGGGATAACTACATA
CTCATGATTATGAAACGCTCACTTTATTTGAAAAACCTCCTAATACACCAAATATGTCACTAGA
TTCCAAAACGTAGACC
AATTGTATCTAATCTCAAATTCTCAATCAAAGTATTAATTTACCGATGGTAAGAAAAGTTAACC
GATATAATTATCAAAA
GAAAGAATAAGTCAACAGATTCTTAATCTCTTTATTTTGGTATATGAACATTTGTACAAAAATC
TCAAAAGATATGTAAC
TGTTTAAAATATAAATTCACTGAGATTAATTCTTCAGACTCGTGTTAGCTATAATAATGTCAAG
AGTTCTTCTTGTTTCA
AGGAAAAACCTTAAAGATATGTATATTTTCTGTAATTATGATGATATAATTTGCTATTCATTGT
CACAAACATTACTTTA
AAAAATCGTATTTTCATTACTACAATGTTGACTAAGAACAAAAATACATTGATTATTGATATAT
CGTCAACTGAATTTTC
TTCCGAGGGATATAATTCTCAAACATAGCAAGAATCTCATAATAATGTTTCGTGACTACCTTTA
GACGAAATTTTTTTAA
GATTCGTAACGTGACTTATGGTCTCTTGCTGTGGGGGTCAATGCGAATAAATCTAAATGTATG
GGAGTCAAATAAAATAC
CAAGAAAAATAAAGGAGCAGCACCCAATAAACTATATGGGACCAGAAATCCTTTCATTGGTTT
AAAATAGGATTATCCCG
AAAGATGAAGGACTAAATTGAAACTGATTGGGGGTAGGAAGAGATCCGTCACAATCATTAAT
GGCTTCCACGCGGAAACT
TGTCGTTTATACAATTTCATTAACTTTCGGGTCGGGTTTATATTCCAAATGGGTCAAATAATAT
TAGTTTAATACACTAA
CGGAGTAATTAATTGGTGACTACAATTTTATCAGTTTGGTGCAATTAGAAACGAACATAGTCG
TAAAATACGAGTTCGGT
GTTATACCTTTATTTACGTTAAAAAAATACGAGAATTTTGTGTCAAATTTCAAATTAATTTCAT
GAATATATGGAAATTA
TTAGATACTCTAGCGAAAATAGTGATTATGAGCGTTTTACAAAAATACGATTTTAGCATTGAA
CTTCCTTTATGTAATTC

GGTCAAATGTTGGCATGAAGAAGCAAGTTTGCAACATTAAATTTCATTTAAAAATCGTGTTGA
CATACTTTAAAATCTAA
ATATAGGAAGAAGACCAAAACATTAAATTTAGTAAGATTCTAATGAACATTTATAAGTTATAA
CTTATAACCAACAAAAG
TTGGGTTTAGCGTTGTTGCTTTATCTGAAAACTTGCAAACTAAACCATTTTAATAGGACTAATG
ACAATTAACAACAAAA
TACACTTAAGCAACAACGTCCTCGTGAATATAATTTGGGCCTCAGGCCCATATTGCTAACGCC
AACTGATATTTCACTTT
ATTCCTTCTTCATCTCACCACACTCTCTCTCTATCTCTATCTCTAACGGCATAGCTGACTCAGTG
TTCTCCGGCATTGAC
TCGCCTGAGAATCAGAAAGCTTAGATCGGTGAGCTTTTAGCTCCATTTTCTGTTTATTTACATA
TTATTTCCTTTTTTTC
TCTCTCCCTTTTTTATCTGGAATTTGTTCTGCTAAATTTTCCAGCTGTTACATTTTCCGATCACG
AGAAGAATCACTGGG
TTTTTATGTTAATCAATACATGTTCCTGTTTTCTGATCATAAATCTCAGCTATTAACACCTGATT
TTGATTCTGCGTAAT
AAAAACCTCTGATTTGCTTTTATCTTCACTTTCCCCATAAACATTGCTTACTTTATTCGCTCTTC
TTTTACCGTTTCCAG
CTAAAAAATTCTTCGCTATTCAATGTGTTTCTCGTTTTGTTGATGAGAAAAATATCTGACAAAA
AATCATTTATTGCATT
TTATGGTGCAGATTCTTAGTTAATGTCGCCTTCTCTAACCAAGTCAGATTAAAAAGGAGTGTTC
GTCCATGTTGCTTTGT
TTTGGTGTTTGGAGAGAGTTTTCGGAGAGTTAGGTGAGTGTTATTTGGGGTGAGGTAGTGATA
AGGTTTGAAGGGGGAGT
GATTCATCAAGTGTGTTATGAATTCGAGGGCTGATCCGGGGGATAGATATTTTCGAGTTCCTTT
GGAGAATCAAACTCAA
CAAGAGTTCATGGGTTCTTGGATTCCATTTACACCCAAAAAACCTAGATCAAGTCTGATGGTA
GATGAGAGAGTGATAAA
CCAGGATCTAAATGGGTTTCCAGGTGGTGAATTTGTAGACAGGGGATTCTGCAACACTGGTGT
GGATCATAATGGGGTTT
TTGATCATGGTGCTCATCAGGGCGTTACCAACTTAAGTATGATGATCAATAGCTTAGCGGGAT
CACATGCACAAGCTTGG
AGTAATAGTGAGAGAGATCTTTTGGGCAGGAGTGAGGTGACTTCTCCTTTAGCACCAGTTATC
AGAAACACCACCGGTAA
TGTAGAGCCGGTCAATGGAAATTTTACTTCAGATGTGGGTATGGTAAATGGTCCTTTCACCCA
GAGTGGCACTTCTCAAG
CTGGCTATAATGAGTTTGAATTGGATGACTTGTTGAATCCTGATCAGATGCCCTTCTCCTTCAC
AAGCTTGCTGAGTGGT
GGGGATAGCTTATTCAAGGTTCGTCAATGTGAGTGATCAAATCTATTTTCAGTTTTTTTTTTTCC
CTTTCTTCCGTTCTT

GCAGTACTTAGAGTAGAACATGAATTAGAATATCTTAAGAAAGTCATGGTTTTGAACAGATGG
ACCTCCAGCGTGTAACA
AGCCTCTTTACAATTTGAATTCACCAATTAGAAGAGAAGCAGTTGGGTCAGTCTGTGAAAGTT
CGTTTCAATATGTACCG
TCAACGCCCAGTCTGTTCAGAACAGGTGAAAAGACTGGATTCCTTGAACAGATAGTTACAACT
ACTGGACATGAAATCCC
AGAGCCGAAATCTGACAAAAGTATGCAGAGCATTATGGACTCGTCTGCTGTTAATGCGACGGA
AGCTACTGAACAAAATG
ATGGCAGCAGACAAGATGTTCTGGAGTTCGACCTTAACAAAACTCCTCAGCAGAAACCCTCCA
AAAGGAAAAGGAAGTTC
ATGCCCAAGGTGGTCGTGGAAGGCAAACCTAAAAGAAAGCCACGCAAACCTGCAGAACTTCC
CAAAGTGGTCGTGGAAGG
CAAACCTAAAAGGAAGCCACGCAAAGCTGCAACTCAGGAAAAAGTGAAATCTAAAGAAACCG
GGAGTGCCAAAAAGAAAA
ATTTGAAAGAATCAGCAACTAAAAAGCCAGCCAATGTTGGAGATATGAGCAACAAAAGCCCT
GAAGTCACACTCAAAAGT
TGCAGAAAAGCTTTGAATTTTGACTTGGAGAATCCTGGAGATGCGAGGCAAGGTGACTCTGAG
TCTGAAATTGTCCAGAA
CAGTAGTGGCGCAAACTCGTTTTCTGAGATCAGAGATGCCATTGGTGGAACTAATGGTAGTTT
CCTGGATTCAGTGTCAC
AAATAGACAAGACCAATGGATTGGGGGCTATGAACCAGCCACTTGAAGTGTCAATGGGAAAC
CAGCCAGATAAACTATCT
ACAGGAGCGAAACTGGCCAGAGACCAACAACCTGATTTATTGACTAGAAACCAGCAATGCCA
GTTCCCAGTGGCAACCCA
GAACACCCAGTTCCCAATGGAAAACCAACAAGCTTGGCTTCAGATGAAAAACCAACTTATTGG
CTTTCCATTTGGTAACC
AGCAACCTCGCATGACCATAAGAAACCAGCAGCCTTGCTTGGCCATGGGTAATCAACAACCTA
TGTATCTGATAGGAACT
CCACGGCCTGCATTAGTAAGTGGAAACCAGCAACTAGGAGGTCCCCAAGGAAACAAGCGGCC
TATATTTTTGAATCACCA
GACTTGTTTACCTGCTGGAAATCAGCTATATGGATCACCTACAGACATGCATCAACTTGTTATG
TCAACCGGAGGGCAAC
AACATGGACTACTGATAAAAAACCAGCAACCTGGATCATTAATAAGAGGCCAGCAGCCTTGC
GTACCTTTGATTGACCAG
CAACCTGCAACTCCAAAAGGTTTTACTCACTTGAATCAGATGGTAGCTACCAGCATGTCATCG
CCTGGGCTTCGACCTCA
TTCTCAGTCACAAGTTCCTACAACATATCTACATGTGGAATCTGTTTCCAGGATTTTGAATGGG
ACTACAGGTACATGCC
AGAGAAGCAGGGCTCCTGCATACGATTCTTTACAGCAAGATATCCATCAAGGAAATAAGTACA
TACTTTCTCATGAGATA

TCCAATGGTAATGGGTGCAAGAAAGCGTTACCTCAAAACTCTTCTCTGCCAACTCCAATTATG
GCTAAACTTGAGGAAGC
CAGGGGCTCGAAGAGACAGTATCATCGTGCAATGGGACAGACGGAAAAGCATGATCTAAACT
TAGCTCAACAGATTGCTC
AATCACAAGATGTGGAGAGACATAACAGCAGCACGTGTGTGGAATATTTAGATGCTGCAAAG
AAAACGAAAATCCAGAAA
GTAGTCCAAGAAAATTTGCATGGCATGCCACCTGAGGTTATAGAAATCGAGGATGATCCAACT
GATGGGGCAAGAAAAGG
TAAAAATACTGCCAGCATCAGTAAAGGTGCATCTAAAGGAAACTCGTCTCCAGTTAAAAAGAC
AGCAGAAAAGGAGAAAT
GTATTGTCCCAAAAACGCCTGCAAAAAAGGGTCGAGCAGGTAGAAAAAAATCAGTACCTCCG
CCTGCTCATGCCTCAGAG
ATCCAGCTTTGGCAACCTACTCCTCCAAAGACACCTTTATCAAGAAGCAAGCCTAAAGGAAAA
GGGAGAAAGTCCATACA
AGATTCAGGAAAAGCAAGAGGTAACTAATGTATTCTACAATCTCTGTGATATAATTTTGAGAT
TTTAGTAACTGATGTGT
CCAAACCAGCTCCTTATCACTGTTGGTGCGTTGTATAGGTCCATCAGGAGAACTTCTGTGTCAG
GATTCTATTGCGGAAA
TAATTTACAGGATGCAAAATCTGTATCTAGGAGACAAAGAAAGAGAACAAGAGCAAAATGCA
ATGGTCTTGTACAAAGGA
GATGGTGCACTTGTTCCCTATGAGAGCAAGAAGCGAAAACCAAGACCCAAAGTTGACATTGA
CGATGAAACAACTCGCAT
ATGGAACTTACTGATGGGGAAAGGAGATGAAAAAGAAGGGGATGAAGAGAAGGATAAAAAG
AAAGAGAAGTGGTGGGAAG
AAGAAAGAAGAGTCTTCCGAGGAAGGGCTGATTCCTTCATCGCTCGCATGCACCTGGTACAAG
GTGAAGATCCACTTCTC
TTCTCAACTCCATTTTTATTCACACAAATTAGTAGAATACTCAAAAATGATGTTTTGTTTGCAA
AATTTTAAAATTCACT
AGTTAACCATGTCAAATAATATTCATAATGCATCTTGTGAAGAACAGGTGTGCATTTATGGTG
ACAGCTGAATGGTTTAT
GTGCCTATTATTTCTTTTACTGCTATAGATGACCAATTGAACTTAAACGTTTACAGGAGATAGA
CGTTTTTCGCCATGGA
AGGGATCGGTGGTTGATTCGGTCATTGGAGTTTTCCTTACACAGAATGTCTCGGATCACCTTTC
AAGGTATATGAGTTGC
CTTAATAAATTGAGTTCCAAAACATAGAAATTAACCCATGGTGGTTTTACAATGCAGCTCTGC
GTTCATGTCTCTAGCTG
CTCGATTCCCTCCAAAATTAAGCAGCAGCCGAGAAGATGAAAGGAATGTTAGAAGCGTAGTT
GTTGAAGATCCAGAAGGA
TGCATTCTGAACTTAAATGAAATTCCTTCGTGGCAGGAAAAGGTTCAACATCCATCTGACATG
GAAGTTTCTGGGGTTGA

TAGTGGATCAAAAGAGCAGCTAAGGGACTGTTCAAACTCTGGAATTGAAAGATTTAATTTCTT
AGAGAAGAGTATTCAAA
ATTTAGAAGAGGAAGTATTATCATCACAAGATTCTTTTGATCCGGCGATATTTCAGTCGTGTGG
GAGAGTTGGATCCTGT
TCATGTTCCAAATCAGACGCAGAGTTTCCTACAACCAGGTGTGAAACAAAAACTGTCAGTGGA
ACATCACAATCAGTGCA
AACTGGGAGCCCAAACTTGTCTGATGAAATTTGTCTTCAAGGGAATGAGAGACCGCATCTATA
TGAAGGATCTGGTGATG
TTCAGAAACAAGAAACTACAAATGTCGCTCAGAAGAAACCTGATCTTGAAAAAACAATGAAT
TGGAAAGACTCTGTCTGT
TTTGGTCAGCCAAGAAATGATACTAATTGGCAAACAACTCCTTCCAGCAGCTATGAGCAGTGT
GCGACTCGACAGCCACA
TGTACTAGACATAGAGGATTTTGGAATGCAGGGTGAAGGCCTTGGTTATTCTTGGATGTCCAT
CTCACCAAGAGTTGACA
GAGTAAAGAACAAAAATGTACCACGCAGGTTTTTCAGACAAGGTGGAAGTGTTCCAAGAGAA
TTCACAGGTCAGATCATA
CCATCAACGCCTCATGAATTACCAGGAATGGGATTGTCCGGTTCCTCAAGCGCCGTCCAAGAA
CACCAGGACGATACCCA
ACATAATCAACAAGATGAGATGAATAAAGCATCCCATTTACAAAAAACATTTTTGGATCTGCT
CAACTCCTCTGAAGAAT
GCCTTACAAGACAGTCCAGTACCAAACAGAACATCACGGATGGCTGTCTACCGAGAGATAGA
ACTGCTGAAGACGTGGTT
GATCCGCTCAGTAACAATTCAAGCTTACAGAACATATTGGTCGAATCAAATTCCAGCAATAAA
GAGCAGACGGCAGTTGA
ATACAAGGAGACAAATGCCACTATTTTACGAGAGATGAAAGGGACGCTTGCTGATGGGAAAA
AGCCTACAAGCCAGTGGG
ATAGTCTCAGAAAAGATGTGGAGGGGAATGAAGGGAGACAGGAACGAAACAAAAACAATAT
GGATTCCATAGACTATGAA
GCAATAAGACGTGCTAGTATCAGCGAGATTTCTGAGGCTATCAAGGAAAGAGGGATGAATAA
CATGTTGGCCGTACGAAT
TAAGGTAAATCTACTAATTTCAGTTGAGACCCTCATCAAATCTGTCAGAAGGCTTGAACATCA
GTAAATTATGTAACCAT
ATTTACAACATTGCAGGATTTCCTAGAACGGATAGTTAAAGATCATGGTGGTATCGACCTTGA
ATGGTTGAGAGAATCTC
CTCCTGATAAAGCCAAGTGGGTAAATCACATTTTTAGTGACTGCAACACTAGCACGATCGATT
TACTCAACAATTACGTC
AAACTGAGTATTAACAAGTTGCTCATGAACATTTCACAGGGACTATCTCTTGAGCATAAGAGG
TCTGGGTTTGAAAAGTG
TTGAATGCGTGCGACTCTTAACACTCCACAATCTTGCTTTCCCTGTGAGTCAGACTATTCCATT
ATCTACTAAAAACTTA

GAATAACTCCGGCTAACTAAGCTGGAACTTGTATTGATGATATGAAGGTTGACACGAATGTTG
GAAGGATAGCAGTTAGG
ATGGGATGGGTGCCTCTACAACCCCTACCTGAATCACTTCAGTTACACCTCCTGGAGCTGTAA
GTTTCTTTTTGTTTGTC
ATCTAAACAACGAAATTTTTATGCAAGTCATAACCATGCTGTGTTTTCACAGATACCCAGTGCT
CGAGTCCATCCAAAAA
TTTCTTTGGCCAAGACTTTGCAAACTCGATCAACGAACACTGTATGCTCATAAACTCTAACAAA
TCATCTGTCTGAAAAA
CCAATATTTCTTTGGTAGAATTCTATTGTCATTACTCATTACTAACAGCGAAATTAATTAACGT
TCTTTTTCTTACTCAG
GTATGAATTACACTACCAACTGATTACGTTTGGAAAGGTATTATTGCTCTAAGCTTTGAATTTA
TCATATGGTAATTTCA
AGCATTGTAGGCACCTGATCAATTATGTGTCTAAATCATGTGAATTCATGTCAGGTATTTTGCA
CAAAGAGTAGACCAAA
TTGTAATGCATGTCCAATGAGAGGAGAGTGCAGACACTTTGCCAGTGCTTATGCTAGGTAAGC
AAGCTTTCATGTACTTA
TATGCAATAATTAAAGATAAAATTTAGGATTATGGGTAAGTTACAAAAAATTAGGCTCAGTTT
CATGGTAGCTAGCTGGA
AATAGTATTACAAGAACAACATAAAGATCAAAGACAGAATCATGGATCCATATGCACTATCAT
TTTAGCTCTTGTAATCC
ATACATGAACACTATATGCCAAAGTAGGGATTTCAAATATGAGATTCGATGACTGATGCCATT
GTAACAGTGCAAGACTT
GCTTTACCGGCACCAGAGGAGAGGAGCTTAACAAGTGCAACTATTCCGGTCCCTCCCGAGTCC
TATCCTCCTGTAGCCAT
CCCGATGATAGAACTACCTCTTCCGTTGGAGAAATCCCTAGCAAGTGGAGCACCATCGAATAG
AGAAAACTGTGAACCAA
TAATTGAAGAGCCGGCCTCGCCCGGGCAAGAGTGCACTGAAATAACCGAGAGTGATATTGAA
GATGCTTACTACAATGAG
GACCCTGACGAGATCCCAACAATAAAACTCAACATTGAACAGTTTGGAATGACTCTACGGGAA
CACATGGAAAGAAACAT
GGAGCTCCAAGAAGGTGACATGTCCAAGGCTTTGGTTGCTTTGCATCCAACAACTACTTCTATT
CCAACTCCCAAACTAA
AGAACATTAGCCGTCTCAGGACAGAGCACCAAGTGTAAGCTAATATCTCCTCCTATATTTTATC
TTCCATATAAATTTTG
GGGAAAAAATCGCTCTCCATCTGGTTTTAGAACATGCGGGTCAGCCAGGGTTATGGCATTTTT
ATATATTTCACCGATCG
GCCCGAGCTGGCTCTGGTTGACTCGTATGCCACCCTGCATTGAACAAACCAGTAGGAGACAAG
CAAGCAAAACGTTTTAA
GATAAGGTCTATGGTAAAATGACAAGGTAACTGATAAATGTGTCGTCTATTTGCAGGTACGAG
CTCCCAGATTCACATCG

TCTCCTTGATGGTGTAAGTCAATTTTTAACTCTCTCTATACTCGAGTTGTTTCACTTGAGCAACA
CTGTTTAAAAGTCCT
CATTTGATAAAATAACAGATGGATAAAAGAGAACCAGATGATCCAAGTCCTTATCTCTTAGCT
ATATGGACACCAGGTGA
GAATAAAACTGCAATGTTTCATTCATGTGTCTACAGTATCAAAGAAAGTACAGCTAGAGCTAA
AAAGCATTTGAAATAGA
GTCGGTTAAATATGAAAGTTTGAATCTGTAAATGAAAGCCGGAACGTAGCATTGGTGGATGTT
ATATGTAAATTAGTTTT
TGAGATTGGTCTAATGTAGTTGTTTGACTGCCAGGTGAAACAGCGAATTCGGCACAACCGCCT
GAACAGAAGTGTGGAGG
GAAAGCGTCTGGCAAAATGTGCTTTGACGAGACTTGTTCTGAGTGTAACAGTCTGAGGGAAGC
AAACTCACAGACAGTTC
GAGGAACTCTTCTGGTGAGATTATCTTGATCTTTTGTGTTGCTCATGAAAAGGAGAAGTGAGA
ATACAAGTTTGCTAATA
TCATTTTTTCGTCATTCACAGATACCTTGTCGGACTGCCATGAGAGGAAGTTTTCCGCTCAACG
GGACATATTTCCAAGT
CAACGAGGTTAGATGAAATAAAACTCAAACAGACAGACGAAACATTATTTCTGTTTAGTGTTG
GTTCTTTATCCTCCTTG
CCATTTTTTATCTTGCAGTTATTTGCAGACCACGAGTCCAGTCTCAAACCCATCGATGTTCCTA
GAGATTGGATATGGGA
TCTCCCAAGAAGGACTGTTTACTTCGGAACATCAGTAACATCAATATTCAGAGGTAAAAACAT
TCGTAATAGAGTTAGTT
AATCAAATGTCCAAAACACAAGAAAGCTTCACCGTCCAATACACAAGAAAGCTTCACCTTCTC
TTTGCCAAAAAAGATCT
TAGAATGTTTTGCTGAATTTGTGCAGGTCTTTCAACGGAGCAGATACAGTTCTGCTTTTGGAAA
GGTAAACGTTAACTTT
CGACCCAGAGAAATCCGGAAAATCTATTGCTTTGTTCTGATCAATACGTTAAACATATACACA
CACACTTTACACTTAGG
ACCAATACTGTTCTGATCTGTGATAGAAACTGGTAAACATCTAACAATTATGATTGCAGGATT
CGTATGTGTCCGTGGAT
TCGAACAGAAGACAAGAGCACCGCGTCCATTAATGGCAAGGTTGCATTTTCCTGCGAGCAAAT
TGAAGAACAACAAAACC
TAAAGATGACTGGAAGAAAGCAAACGCATTGCTTCTCTGCTCTCCTCTATTTAAAGCCAGGAA
AAGTCCCATTTAGACAT
AATAACAGGAATCCAAATAGGCTATTTTCTCTTTCTTTCTTATTTCATTCATAGAGCAGAAGCG
ACACAAAAAAAGTTTTT
TGGGTTATTTATTTTCTCTCTAACAAATTTGTAGCGTTTTGGGTCTTTTTCTGGCTGTCACTAGC
GTGGCAAATCCAATG
TCCGCGCACACTTAGGCGCATTGTCAATAAATTCTCCGGCCACCGGAGTGTTACGATCTTTTCC
AACGGCGGCTAATGCG

ATATTTCCGGTAACACATATTCCTTATTCTATGTTGGTTTTGTGTACGGCGTGGGCCTTACTAG
ACAATGATCATCAATA
AAACTAACACAAAGTTGAATGCTACAAAGTAGAAAGTGAAGAAAAAATAATATAGACATTGC
CGA

SEQ ID NO:2 DMT amino acid sequence

MQSIMDSSAVNATEATEQNDGSRQDVLEFDLNKTPQQKPSKRKRKFMPKVVVEGKPKRKPRKPA
ELPKVVVEGKPKRKPR
KAATQEKVKSKETGSAKKKNLKESATKKPANVGDMSNKSPEVTLKSCRKALNFDLENPGDARQG
DSESEIVQNSSGANSF
SEIRDAIGGTNGSFLDSVSQIDKTNGLGAMNQPLEVSMGNQPDKLSTGAKLARDQQPDLLTRNQQ
CQFPVATQNTQFPME
NQQAWLQMKNQLIGFPFGNQQPRMTIRNQQPCLAMGNQQPMYLIGTPRPALVSGNQQLGGPQGN
KRPIFLNHQTCLPAGN
QLYGSPTDMHQLVMSTGGQQHGLLIKNQQPGSLIRGQQPCVPLIDQQPATPKGFTHLNQMVATSM
SSPGLRPHSQSQVPT
TYLHVESVSRILNGTTGTCQRSRAPAYDSLQQDIHQGNKYILSHEISNGNGCKKALPQNSSLPTPIM
AKLEEARGSKRQY
HRAMGQTEKHDLNLAQQIAQSQDVERHNSSTCVEYLDAAKKTKIQKVVQENLHGMPPEVIEIEDD
PTDGARKGKNTASIS
KGASKGNSSPVKKTAEKEKCIVPKTPAKKGRAGRKKSVPPPAHASEIQLWQPTPPKTPLSRSKPKG
KGRKSIQDSGKARG
PSGELLCQDSIAEIIYRMQNLYLGDKEREQEQNAMVLYKGDGALVPYESKKRKPRPKVDIDDETTR
IWNLLMGKGDEKEG
DEEKDKKKEKWWEEERRVFRGRADSFIARMHLVQGDRRFSPWKGSVVDSVIGVFLTQNVSDHLS
SSAFMSLAARFPPKLS
SSREDERNVRSVVVEDPEGCILNLNEIPSWQEKVQHPSDMEVSGVDSGSKEQLRDCSNSGIERFNFL
EKSIQNLEEEVLS
SQDSFDPAIFQSCGRVGSCSCSKSDAEFPTTRCETKTVSGTSQSVQTGSPNLSDEICLQGNERPHLYE
GSGDVQKQETTN
VAQKKPDLEKTMNWKDSVCFGQPRNDTNWQTTPSSSYEQCATRQPHVLDIEDFGMQGEGLGYS
WMSISPRVDRVKNKNVP
RRFFRQGGSVPREFTGQIIPSTPHELPGMGLSGSSSAVQEHQDDTQHNQQDEMNKASHLQKTFLDL
LNSSEECLTRQSST
KQNITDGCLPRDRTAEDVVDPLSNNSSLQNILVESNSSNKEQTAVEYKETNATILREMKGTLADGK
KPTSQWDSLRKDVE
GNEGRQERNKNNMDSIDYEAIRRASISEISEAIKERGMNNMLAVRIKDFLERIVKDHGGIDLEWLRE
SPPDKAKDYLLSI
RGLGLKSVECVRLLTLHNLAFPVDTNVGRIAVRMGWVPLQPLPESLQLHLLELYPVLESIQKFLWP
RLCKLDQRTLYELH

YQLITFGKVFCTKSRPNCNACPMRGECRHFASAYASARLALPAPEERSLTSATIPVPPESFPPVAIPM
IELPLPLEKSLA
SGAPSNRENCEPIIEEPASPGQECTEITESDIEDAYYNEDPDEIPTIKLNIEQFGMTLREHMERNMELQ
EGDMSKALVAL
HPTTTSIPTPKLKNISRLRTEHQVYELPDSHRLLDGMDKREPDDPSPYLLAIWTPGETANSAQPPEQ
KCGGKASGKMCFD
ETCSECNSLREANSQTVRGTLLIPCRTAMRGSFPLNGTYFQVNELFADHESSLKPIDVPRDWIWDLP
RRTVYFGTSVTSI
FRGLSTEQIQFCFWKGFVCVRGFEQKTRAPRPLMARLHFPASKLKNNKT

SEQ ID NO:3 DMT 5' flanking sequence

AAGCTTAAAGCTACCAACATCGAATTTAGTAAAAGACCCATGATTTGAAATTGGAATTGTCGG
CAAAATCGAGAAGATAT
AGAGCCGACACGGGAACAGTGAAAACCACAAAGCGCGTAAGAATGAAACAGTGGGAGAAGG
AAGAGAGAATCTTACCGAT
CATTCGAGGGAAAAGATGGGAATCAGAGAAAAATCTGGAAAAAAAGAAATTAAGAGAAAGA
GAGAGAAGAAAGTGAGGAG
GAAGATGCAGTGAAGACTGCTATAGCCACATCCCACATGGTGTGATGAGAGAGAGAGAGAGA
GAGGTTAAAGCAGCAAAT
TGTGGAGAGATAAAGAGAGAGAGAGACTGAGCGAGTCAAGTTCGTCGTCGTGTTTAAAAGAA
AGAATCCTATATTTGCCT
TTTTCTTTACTACTTTATTTTCAGACTATTTGCTTATTTTGCCTCAAACTTTTTTGATTGTCACTT
TTCGATCCTAAAGT
GTTTGACAATTTACCTGCCTTTTTCTCCAAGAAAAATCAGAACAGACCACAGCAAATTTATGTA
TTTTCTATTAAAAAAG
AAAGAAAGAATTCATATTACTTATAGAATTAAAAGCTAAGCAGTTGAAAACGTGAAAGCAGA
ATTTCTAAAAAAAAATAGT
AAACTGCTACAAACTTATTTATGTGTATATAACATATCTATAAAGAAACTCAAATATATGATA
AATCATTTTAACAAAAT
TTCTATGAAATTATAATAAAAAAAGTCACTTTTGACACTTAAAAGGTTGACAATAACCGTCTCT
CCAAAAAAAAAATCAAA
ACATTTATAATTTCTAAAACTATGGTGTAATTTTGCTGAAATCAAAAAGAAAAGAAGGATTTC
TATATCATAAGTTTCAT
TATTGTATCAAACTTTCAAATTTCATGTAATTTGAAAGGAAAAAAATTAAGATATAATGTTGTT
TTTGTTTCTTATGTTA
CATTTTCATGGAATATATATTCATAACAAAAAATGTATTTTAATATGATGAGAGATTACCATCC
AAAAGGTCGAACTTAT
ATAAAACAAGTTAATAACTAAACAATACATGTGATCACAATCAATGACAGTTTTGATCTTAAA
ATAGAAATGATTGAGCA

AACCTCAAAAATGTCTTCTTAGGATCACAAAATCTTTCCTTTAGCTTATTAAAGCCGGGAGTTC
AACTCTCTCTCCCTTG
TAGACTTTTTGTTTTCAAATCTTTTTCTTTCAAAAAATCAATAATTAGTTAATGGGCATAATATT
TGGTTTTAATTAAGT
CCATAGATTTTTTAGGACCATCTCTAATCACGACAAATATCCTAAATTGTAACACATTTAAAAC
TTAAAAGTATTGCATT
CACAATCCTTAAAATATATATATATATATATATATATATATATATATATATATATATGAAAGTTAT
ATAGAAACGATAACTC
CTTACTCAACAATTAGCCCAAAAAAACATCCATAATGCATTTAAACTAGGAATTTTAACAAAC
TCAAATAGGTTGGTAGT
TAAAAAAAAACAAATAGTAGATGTACATACGTACCTTTAAAAATATATACTCATATCGAAAGT
TTTAAATTTTGCGAAAT
TAAATACATTTATCTATCAATTAAAATACATTTAATAATGCATAATTCTGTAATATCTATCTTT
AATTTCCATATAGAAC
CAAAACAAAATAAACATATCAAATAGTTTTAACTTAACAAAAACGTTAGGGAAAAGTTGACCT
AACTAGCTTGATTGACG
TTGAACTTGTCAATGCGAAAGCGATATTTCCAATATATACTACATGTAGTATTATTTATATGGA
AGTTTCTAAAAAGGTG
TTGAGTGGATTGTTACTTGTTGGAGGATGCTATTTTTTCCTTCTTGCCATAATATTTTACGAGTA
TGGGATAACTACATA
CTCATGATTATGAAACGCTCACTTTATTTGAAAAACCTCCTAATACACCAAATATGTCACTAGA
TTCCAAAACGTAGACC
AATTGTATCTAATCTCAAATTCTCAATCAAAGTATTAATTTACCGATGGTAAGAAAAGTTAACC
GATATAATTATCAAAA
GAAAGAATAAGTCAACAGATTCTTAATCTCTTTATTTTGGTATATGAACATTTGTACAAAAATC
TCAAAAGATATGTAAC
TGTTTAAAATATAAATTCACTGAGATTAATTCTTCAGACTCGTGTTAGCTATAATAATGTCAAG
AGTTCTTCTTGTTTCA
AGGAAAAACCTTAAAGATATGTATATTTTCTGTAATTATGATGATATAATTTGCTATTCATTGT
CACAAACATTACTTTA
AAAAATCGTATTTTCATTACTACAATGTTGACTAAGAACAAAAATACATTGATTATTGATATAT
CGTCAACTGAATTTTC
TTCCGAGGGATATAATTCTCAAACATAGCAAGAATCTCATAATAATGTTTCGTGACTACCTTTA
GACGAAATTTTTTTAA
GATTCGTAACGTGACTTATGGTCTCTTGCTGTGGGGGTCAATGCGAATAAATCTAAATGTATG
GGAGTCAAATAAAATAC
CAAGAAAAATAAAGGAGCAGCACCCAATAAACTATATGGGACCAGAAATCCTTTCATTGGTTT
AAAATAGGATTATCCCG
AAAGATGAAGGACTAAATTGAAACTGATTGGGGGTAGGAAGAGATCCGTCACAATCATTAAT
GGCTTCCACGCGGAAACT

TGTCGTTTATACAATTTCATTAACTTTCGGGTCGGGTTTATATTCCAAATGGGTCAAATAATAT
TAGTTTAATACACTAA
CGGAGTAATTAATTGGTGACTACAATTTTATCAGTTTGGTGCAATTAGAAACGAACATAGTCG
TAAAATACGAGTTCGGT
GTTATACCTTTATTTACGTTAAAAAAATACGAGAATTTTGTGTCAAATTTCAAATTAATTTCAT
GAATATATGGAAATTA
TTAGATACTCTAGCGAAAATAGTGATTATGAGCGTTTTACAAAAATACGATTTTAGCATTGAA
CTTCCTTTATGTAATTC
GGTCAAATGTTGGCATGAAGAAGCAAGTTTGCAACATTAAATTTCATTTAAAAATCGTGTTGA
CATACTTTAAAATCTAA
ATATAGGAAGAAGACCAAAACATTAAATTTAGTAAGATTCTAATGAACATTTATAAGTTATAA
CTTATAACCAACAAAAG
TTGGGTTTAGCGTTGTTGCTTTATCTGAAAACTTGCAAACTAAACCATTTTAATAGGACTAATG
ACAATTAACAACAAAA
TACACTTAAGCAACAACGTCCTCGTGAATATAATTTGGGCCTCAGGCCCATATTGCTAACGCC
AACTGATATTTCACTTT
ATTCCTTCTTCATCTCACCACACTCTCTCTCTATCTCTATCTCTAACGGCATAGCTGACTCAGT

SEQ ID NO:4 DMT 3' flanking sequence

AGATGACTGGAAGAAAGCAAACGCATTGCTTCTCTGCTCTCCTCTATTTAAAGCCAGGAAAAG
TCCCATTTAGACATAAT
AACAGGAATCCAAATAGGCTATTTTCTCTTTCTTTCTTATTTCATTCATAGAGCAGAAGCGACA
CAAAAAAGTTTTTTGG
GTTATTTATTTTCTCTCTAACAAAAAAAAAAAAAAAAAACTCGAG

SEQ ID NO:5 DMT cDNA sequence

GTTCTCCGGCATTGACTCGCCTGAGAATCAGAAAGCTTAGATCGGTGAGCTTTTAGCTCCATTT
TCTGTTTATTTACATA
TTATTTCCTTTTTTTCTCTCTCCCTTTTTTATCTGGAATTTGTTCTGCTAAATTTTCCAGCTGTTA
CATTTTCCGATCAC
GAGAAGAATCACTGGGTTTTTATGTTAATCAATACATGTTCCTGTTTTCTGATCATAAATCTCA
GCTATTAACACCTGAT
TTTGATTCTGCGTAATAAAAACCTCTGATTTGCTTTTATCTTCACTTTCCCCATAAACATTGCTT
ACTTTATTCGCTCTT
CTTTTACCGTTTCCAGCTAAAAAATTCTTCGCTATTCAATGTGTTTCTCGTTTTGTTGATGAGAA
AAATATCTGACAAAA
AATCATTTATTGCATTTTATGGTGCAGATTCTTAGTTAATGTCGCCTTCTCTAACCAAGTCAGA
TTAAAAAGGAGTGTTC

GTCCATGTTGCTTTGTTTTGGTGTTTGGAGAGAGTTTTCGGAGAGTTAGGTGAGTGTTATTTGG
GGTGAGGTAGTGATAA
GGTTTGAAGGGGGAGTGATTCATCAAGTGTGTTATGAATTCGAGGGCTGATCCGGGGGATAGA
TATTTTCGAGTTCCTTT
GGAGAATCAAACTCAACAAGAGTTCATGGGTTCTTGGATTCCATTTACACCCAAAAAACCTAG
ATCAAGTCTGATGGTAG
ATGAGAGAGTGATAAACCAGGATCTAAATGGGTTTCCAGGTGGTGAATTTGTAGACAGGGGA
TTCTGCAACACTGGTGTG
GATCATAATGGGGTTTTTGATCATGGTGCTCATCAGGGCGTTACCAACTTAAGTATGATGATCA
ATAGCTTAGCGGGATC
ACATGCACAAGCTTGGAGTAATAGTGAGAGAGATCTTTTGGGCAGGAGTGAGGTGACTTCTCC
TTTAGCACCAGTTATCA
GAAACACCACCGGTAATGTAGAGCCGGTCAATGGAAATTTTACTTCAGATGTGGGTATGGTAA
ATGGTCCTTTCACCCAG
AGTGGCACTTCTCAAGCTGGCTATAATGAGTTTGAATTGGATGACTTGTTGAATCCTGATCAGA
TGCCCTTCTCCTTCAC
AAGCTTGCTGAGTGGTGGGGATAGCTTATTCAAGGTTCGTCAATGTGAGTGATCAAATCTATTT
TCAGTTTTTTTTTTTC
CCTTTCTTCCGTTCTTGCAGTACTTAGAGTAGAACATGAATTAGAATATCTTAAGAAAGTCATG
GTTTTGAACAGATGGA
CCTCCAGCGTGTAACAAGCCTCTTTACAATTTGAATTCACCAATTAGAAGAGAAGCAGTTGGG
TCAGTCTGTGAAAGTTC
GTTTCAATATGTACCGTCAACGCCCAGTCTGTTCAGAACAGGTGAAAAGACTGGATTCCTTGA
ACAGATAGTTACAACTA
CTGGACATGAAATCCCAGAGCCGAAATCTGACAAAAGTATGCAGAGCATTATGGACTCGTCTG
CTGTTAATGCGACGGAA
GCTACTGAACAAAATGATGGCAGCAGACAAGATGTTCTGGAGTTCGACCTTAACAAAACTCCT
CAGCAGAAACCCTCCAA
AAGGAAAAGGAAGTTCATGCCCAAGGTGGTCGTGGAAGGCAAACCTAAAAGAAAGCCACGCA
AACCTGCAGAACTTCCCA
AAGTGGTCGTGGAAGGCAAACCTAAAAGGAAGCCACGCAAAGCTGCAACTCAGGAAAAAGTG
AAATCTAAAGAAACCGGG
AGTGCCAAAAAGAAAAATTTGAAAGAATCAGCAACTAAAAAGCCAGCCAATGTTGGAGATAT
GAGCAACAAAAGCCCTGA
AGTCACACTCAAAAGTTGCAGAAAAGCTTTGAATTTTGACTTGGAGAATCCTGGAGATGCGAG
GCAAGGTGACTCTGAGT
CTGAAATTGTCCAGAACAGTAGTGGCGCAAACTCGTTTTCTGAGATCAGAGATGCCATTGGTG
GAACTAATGGTAGTTTC
CTGGATTCAGTGTCACAAATAGACAAGACCAATGGATTGGGGGCTATGAACCAGCCACTTGAA
GTGTCAATGGGAAACCA

GCCAGATAAACTATCTACAGGAGCGAAACTGGCCAGAGACCAACAACCTGATTTATTGACTAG
AAACCAGCAATGCCAGT
TCCCAGTGGCAACCCAGAACACCCAGTTCCCAATGGAAAACCAACAAGCTTGGCTTCAGATGA
AAAACCAACTTATTGGC
TTTCCATTTGGTAACCAGCAACCTCGCATGACCATAAGAAACCAGCAGCCTTGCTTGGCCATG
GGTAATCAACAACCTAT
GTATCTGATAGGAACTCCACGGCCTGCATTAGTAAGTGGAAACCAGCAACTAGGAGGTCCCCA
AGGAAACAAGCGGCCTA
TATTTTTGAATCACCAGACTTGTTTACCTGCTGGAAATCAGCTATATGGATCACCTACAGACAT
GCATCAACTTGTTATG
TCAACCGGAGGGCAACAACATGGACTACTGATAAAAAACCAGCAACCTGGATCATTAATAAG
AGGCCAGCAGCCTTGCGT
ACCTTTGATTGACCAGCAACCTGCAACTCCAAAAGGTTTTACTCACTTGAATCAGATGGTAGCT
ACCAGCATGTCATCGC
CTGGGCTTCGACCTCATTCTCAGTCACAAGTTCCTACAACATATCTACATGTGGAATCTGTTTC
CAGGATTTTGAATGGG
ACTACAGGTACATGCCAGAGAAGCAGGGCTCCTGCATACGATTCTTTACAGCAAGATATCCAT
CAAGGAAATAAGTACAT
ACTTTCTCATGAGATATCCAATGGTAATGGGTGCAAGAAAGCGTTACCTCAAAACTCTTCTCTG
CCAACTCCAATTATGG
CTAAACTTGAGGAAGCCAGGGGCTCGAAGAGACAGTATCATCGTGCAATGGGACAGACGGAA
AAGCATGATCTAAACTTA
GCTCAACAGATTGCTCAATCACAAGATGTGGAGAGACATAACAGCAGCACGTGTGTGGAATA
TTTAGATGCTGCAAAGAA
AACGAAAATCCAGAAAGTAGTCCAAGAAAATTTGCATGGCATGCCACCTGAGGTTATAGAAA
TCGAGGATGATCCAACTG
ATGGGGCAAGAAAAGGTAAAAATACTGCCAGCATCAGTAAAGGTGCATCTAAAGGAAACTCG
TCTCCAGTTAAAAAGACA
GCAGAAAAGGAGAAATGTATTGTCCCAAAAACGCCTGCAAAAAAGGGTCGAGCAGGTAGAAA
AAAATCAGTACCTCCGCC
TGCTCATGCCTCAGAGATCCAGCTTTGGCAACCTACTCCTCCAAAGACACCTTTATCAAGAAG
CAAGCCTAAAGGAAAAG
GGAGAAAGTCCATACAAGATTCAGGAAAAGCAAGAGGTCCATCAGGAGAACTTCTGTGTCAG
GATTCTATTGCGGAAATA
ATTTACAGGATGCAAAATCTGTATCTAGGAGACAAAGAAAGAGAACAAGAGCAAAATGCAAT
GGTCTTGTACAAAGGAGA
TGGTGCACTTGTTCCCTATGAGAGCAAGAAGCGAAAACCAAGACCCAAAGTTGACATTGACG
ATGAAACAACTCGCATAT
GGAACTTACTGATGGGGAAAGGAGATGAAAAAGAAGGGGATGAAGAGAAGGATAAAAAGAA
AGAGAAGTGGTGGGAAGAA

GAAAGAAGAGTCTTCCGAGGAAGGGCTGATTCCTTCATCGCTCGCATGCACCTGGTACAAGGA
GATAGACGTTTTTCGCC
ATGGAAGGGATCGGTGGTTGATTCGGTCATTGGAGTTTTCCTTACACAGAATGTCTCGGATCA
CCTTTCAAGCTCTGCGT
TCATGTCTCTAGCTGCTCGATTCCCTCCAAAATTAAGCAGCAGCCGAGAAGATGAAAGGAATG
TTAGAAGCGTAGTTGTT
GAAGATCCAGAAGGATGCATTCTGAACTTAAATGAAATTCCTTCGTGGCAGGAAAAGGTTCAA
CATCCATCTGACATGGA
AGTTTCTGGGGTTGATAGTGGATCAAAAGAGCAGCTAAGGGACTGTTCAAACTCTGGAATTGA
AAGATTTAATTTCTTAG
AGAAGAGTATTCAAAATTTAGAAGAGGAAGTATTATCATCACAAGATTCTTTTGATCCGGCGA
TATTTCAGTCGTGTGGG
AGAGTTGGATCCTGTTCATGTTCCAAATCAGACGCAGAGTTTCCTACAACCAGGTGTGAAACA
AAAACTGTCAGTGGAAC
ATCACAATCAGTGCAAACTGGGAGCCCAAACTTGTCTGATGAAATTTGTCTTCAAGGGAATGA
GAGACCGCATCTATATG
AAGGATCTGGTGATGTTCAGAAACAAGAAACTACAAATGTCGCTCAGAAGAAACCTGATCTTG
AAAAAACAATGAATTGG
AAAGACTCTGTCTGTTTTGGTCAGCCAAGAAATGATACTAATTGGCAAACAACTCCTTCCAGC
AGCTATGAGCAGTGTGC
GACTCGACAGCCACATGTACTAGACATAGAGGATTTTGGAATGCAAGGTGAAGGCCTTGGTTA
TTCTTGGATGTCCATCT
CACCAAGAGTTGACAGAGTAAAGAACAAAAATGTACCACGCAGGTTTTTCAGACAAGGTGGA
AGTGTTCCAAGAGAATTC
ACAGGTCAGATCATACCATCAACGCCTCATGAATTACCAGGAATGGGATTGTCCGGTTCCTCA
AGCGCCGTCCAAGAACA
CCAGGACGATACCCAACATAATCAACAAGATGAGATGAATAAAGCATCCCATTTACAAAAAA
CATTTTTGGATCTGCTCA
ACTCCTCTGAAGAATGCCTTACAAGACAGTCCAGTACCAAACAGAACATCACGGATGGCTGTC
TACCGAGAGATAGAACT
GCTGAAGACGTGGTTGATCCGCTCAGTAACAATTCAAGCTTACAGAACATATTGGTCGAATCA
AATTCCAGCAATAAAGA
GCAGACGGCAGTTGAATACAAGGAGACAAATGCCACTATTTTACGAGAGATGAAAGGGACGC
TTGCTGATGGGAAAAAGC
CTACAAGCCAGTGGGATAGTCTCAGAAAAGATGTGGAGGGGAATGAAGGGAGACAGGAACG
AAACAAAAACAATATGGAT
TCCATAGACTATGAAGCAATAAGACGTGCTAGTATCAGCGAGATTTCTGAGGCTATCAAGGAA
AGAGGGATGAATAACAT
GTTGGCCGTACGAATTAAGGATTTCCTAGAACGGATAGTTAAAGATCATGGTGGTATCGACCT
TGAATGGTTGAGAGAAT

CTCCTCCTGATAAAGCCAAGGACTATCTCTTGAGCATAAGAGGTCTGGGTTTGAAAAGTGTTG
AATGCGTGCGACTCTTA
ACACTCCACAATCTTGCTTTCCCTGTTGACACGAATGTTGGAAGGATAGCAGTTAGGATGGGA
TGGGTGCCTCTACAACC
CCTACCTGAATCACTTCAGTTACACCTCCTGGAGCTATACCCAGTGCTCGAGTCCATCCAAAAA
TTTCTTTGGCCAAGAC
TTTGCAAACTCGATCAACGAACACTGTATGAATTACACTACCAACTGATTACGTTTGGAAAGG
TATTTTGCACAAAGAGT
AGACCAAATTGTAATGCATGTCCAATGAGAGGAGAGTGCAGACACTTTGCCAGTGCTTATGCT
AGTGCAAGACTTGCTTT
ACCGGCACCAGAGGAGAGGAGCTTAACAAGTGCAACTATTCCGGTCCCTCCCGAGTCCTTTCC
TCCTGTAGCCATCCCGA
TGATAGAACTACCTCTTCCGTTGGAGAAATCCCTAGCAAGTGGAGCACCATCGAATAGAGAAA
ACTGTGAACCAATAATT
GAAGAGCCGGCCTCGCCCGGGCAAGAGTGCACTGAAATAACCGAGAGTGATATTGAAGATGC
TTACTACAATGAGGACCC
TGACGAGATCCCAACAATAAAACTCAACATTGAACAGTTTGGAATGACTCTACGGGAACACAT
GGAAAGAAACATGGAGC
TCCAAGAAGGTGACATGTCCAAGGCTTTGGTTGCTTTGCATCCAACAACTACTTCTATTCCAAC
TCCCAAACTAAAGAAC
ATTAGCCGTCTCAGGACAGAGCACCAAGTGTACGAGCTCCCAGATTCACATCGTCTCCTTGAT
GGTATGGATAAAAGAGA
ACCAGATGATCCAAGTCCTTATCTCTTAGCTATATGGACACCAGGTGAAACAGCGAATTCGGC
ACAACCGCCTGAACAGA
AGTGTGGAGGGAAAGCGTCTGGCAAAATGTGCTTTGACGAGACTTGTTCTGAGTGTAACAGTC
TGAGGGAAGCAAACTCA
CAGACAGTTCGAGGAACTCTTCTGATACCTTGTCGGACTGCCATGAGAGGAAGTTTTCCGCTC
AACGGGACATATTTCCA
AGTCAACGAGTTATTTGCAGACCACGAGTCCAGTCTCAAACCCATCGATGTTCCTAGAGATTG
GATATGGGATCTCCCAA
GAAGGACTGTTTACTTCGGAACATCAGTAACATCAATATTCAGAGGTCTTTCAACGGAGCAGA
TACAGTTCTGCTTTTGG
AAAGGATTCGTATGTGTCCGTGGATTCGAACAGAAGACAAGAGCACCGCGTCCATTAATGGCA
AGGTTGCATTTTCCTGC
GAGCAAATTGAAGAACAACAAAACCTAAAGATGACTGGAAGAAAGCAAACGCATTGCTTCTC
TGCTCTCCTCTATTTAAA
GCCAGGAAAAGTCCCATTTAGACATAATAACAGGAATCCAAATAGGCTATTTTCTCTTTCTTTC
TTATTTCATTCATAGA
GCAGAAGCGACACAAAAAAGTTTTTTGGGTTATTTATTTTCTCTCTAACAAAAAAAAAAAAAA
AAAACTCGAG

SEQ ID NO:6 5' untranslated region of DMT

GTTCTCCGGCATTGACTCGCCTGAGAATCAGAAAGCTTAGATCGGTGAGCTTTTAGCTCCATTT
TCTGTTTATTTACATA
TTATTTCCTTTTTTTCTCTCTCCCTTTTTTATCTGGAATTTGTTCTGCTAAATTTTCCAGCTGTTA
CATTTTCCGATCAC
GAGAAGAATCACTGGGTTTTTATGTTAATCAATACATGTTCCTGTTTTCTGATCATAAATCTCA
GCTATTAACACCTGAT
TTTGATTCTGCGTAATAAAAACCTCTGATTTGCTTTTATCTTCACTTTCCCCATAAACATTGCTT
ACTTTATTCGCTCTT
CTTTTACCGTTTCCAGCTAAAAAATTCTTCGCTATTCAATGTGTTTCTCGTTTTGTTGATGAGAA
AAATATCTGACAAAA
AATCATTTATTGCATTTTATGGTGCAGATTCTTAGTTAATGTCGCCTTCTCTAACCAAGTCAGA
TTAAAAAGGAGTGTTC
GTCCATGTTGCTTTGTTTTGGTGTTTGGAGAGAGTTTTCGGAGAGTTAGGTGAGTGTTATTTGG
GGTGAGGTAGTGATAA
GGTTTGAAGGGGGAGTGATTCATCAAGTGTGTTATGAATTCGAGGGCTGATCCGGGGGATAGA
TATTTTCGAGTTCCTTT
GGAGAATCAAACTCAACAAGAGTTCATGGGTTCTTGGATTCCATTTACACCCAAAAAACCTAG
ATCAAGTCTGATGGTAG
ATGAGAGAGTGATAAACCAGGATCTAAATGGGTTTCCAGGTGGTGAATTTGTAGACAGGGGA
TTCTGCAACACTGGTGTG
GATCATAATGGGGTTTTTGATCATGGTGCTCATCAGGGCGTTACCAACTTAAGTATGATGATCA
ATAGCTTAGCGGGATC
ACATGCACAAGCTTGGAGTAATAGTGAGAGAGATCTTTTGGGCAGGAGTGAGGTGACTTCTCC
TTTAGCACCAGTTATCA
GAAACACCACCGGTAATGTAGAGCCGGTCAATGGAAATTTTACTTCAGATGTGGGTATGGTAA
ATGGTCCTTTCACCCAG
AGTGGCACTTCTCAAGCTGGCTATAATGAGTTTGAATTGGATGACTTGTTGAATCCTGATCAGA
TGCCCTTCTCCTTCAC
AAGCTTGCTGAGTGGTGGGGATAGCTTATTCAAGGTTCGTCAATGTGAGTGATCAAATCTATTT
TCAGTTTTTTTTTTTC
CCTTTCTTCCGTTCTTGCAGTACTTAGAGTAGAACATGAATTAGAATATCTTAAGAAAGTCATG
GTTTTGAACAGATGGA
CCTCCAGCGTGTAACAAGCCTCTTTACAATTTGAATTCACCAATTAGAAGAGAAGCAGTTGGG
TCAGTCTGTGAAAGTTC
GTTTCAATATGTACCGTCAACGCCCAGTCTGTTCAGAACAGGTGAAAAGACTGGATTCCTTGA
ACAGATAGTTACAACTA
CTGGACATGAAATCCCAGAGCCGAAATCTGACAAAAGT

SEQ ID NO:7 >Arabidopsis thaliana DMT1(1DMT5) gene sequence from BAC T32M21 (gi 7406444);

```
64441 tcactagatt ccaaaacgta gaccaattgt atctaatctc aaattctcaa tcaaagtatt
64501 aatttaccga tggtaagaaa agttaaccga tataattatc aaaagaaaga ataagtcaac
64561 agattcttaa tctctttatt ttggtatatg aacatttgta caaaaatctc aaaagatatg
64621 taactgttta aaatataaat tcactgagat taattcttca gactcgtgtt agctataata
64681 atgtcaagag ttcttcttgt ttcaaggaaa aaccttaaag atatgtatat tttctgtaat
64741 tatgatgata taatttgcta ttcattgtca caaacattac tttaaaaaat cgtattttca
64801 ttactacaat gttgactaag aacaaaaata cattgattat tgatatatcg tcaactgaat
64861 tttcttccga gggatataat tctcaaacat agcaagaatc tcataataat gtttcgtgac
64921 tacctttaga cgaaattttt ttaagattcg taacgtgact tatggtctct tgctgtgggg
64981 gtcaatgcga ataaatctaa atgtatggga gtcaaataaa ataccaagaa aaataaagga
65041 gcagcaccca ataaactata tgggaccaga aatcctttca ttggtttaaa ataggattat
65101 cccgaaagat gaaggactaa attgaaactg attgggggta ggaagagatc cgtcacaatc
65161 attaatggct tccacgcgga aacttgtcgt ttatacaatt tcattaactt tcgggtcggg
65221 tttatattcc aaatgggtca aataatatta gtttaataca ctaacggagt aattaattgg
65281 tgactacaat tttatcagtt tggtgcaatt agaaacgaac atagtcgtaa aatacgagtt
65341 cggtgttata cctttattta cgttaaaaaa atacgagaat tttgtgtcaa atttcaaatt
65401 aatttcatga atatatggaa attattagat actctagcga aaatagtgat tatgagcgtt
65461 ttacaaaaat acgattttag cattgaactt cctttatgta attcggtcaa atgttggcat
65521 gaagaagcaa gtttgcaaca ttaaatttca tttaaaaatc gtgttgacat actttaaaat
65581 ctaaatatag gaagaagacc aaaacattaa atttagtaag attctaatga acatttataa
65641 gttataactt ataaccaaca aaagttgggt ttagcgttgt tgctttatct gaaaacttgc
65701 aaactaaacc attttaatag gactaatgac aattaacaac aaaatacact taagcaacaa
65761 cgtcctcgtg aatataattt gggcctcagg cccatattgc taacgccaac tgatatttca
65821 ctttattcct tcttcatctc accacactct ctctctatct ctatctctaa cggcatagct
65881 gactcagtgt tctccggcat tgactcgcct gagaatcaga aagcttagat cggtgagctt
65941 ttagctccat tttctgttta tttacatatt atttcctttt tttctctctc ccttttttat
66001 ctggaatttg ttctgctaaa ttttccagct gttacatttt ccgatcacga gaagaatcac
66061 tgggttttta tgttaatcaa tacatgttcc tgttttctga tcataaatct cagctattaa
66121 cacctgattt tgattctgcg taataaaaac ctctgatttg cttttatctt cactttcccc
66181 ataaacattg cttactttat tcgctcttct tttaccgttt ccagctaaaa aattcttcgc
66241 tattcaatgt gtttctcgtt ttgttgatga gaaaaatatc tgacaaaaaa tcatttattg
66301 cattttatgg tgcagattct tagttaatgt cgccttctct aaccaagtca gattaaaaag
66361 gagtgttcgt ccatgttgct ttgttttggt gtttggagag agttttcgga gagttaggtg
66421 agtgttattt ggggtgaggt agtgataagg tttgaagggg gagtgattca tcaagtgtgt
66481 tatgaattcg agggctgatc cgggggatag atattttcga gttcctttgg agaatcaaac
66541 tcaacaagag ttcatgggtt cttggattcc atttacaccc aaaaaaccta gatcaagtct
66601 gatggtagat gagagagtga taaaccagga tctaaatggg tttccaggtg gtgaatttgt
66661 agacagggga ttctgcaaca ctggtgtgga tcataatggg gtttttgatc atggtgctca
66721 tcagggcgtt accaacttaa gtatgatgat caatagctta gcgggatcac atgcacaagc
66781 ttggagtaat agtgagagag atcttttggg caggagtgag gtgacttctc ctttagcacc
66841 agttatcaga aacaccaccg gtaatgtaga gccggtcaat ggaaatttta cttcagatgt
66901 gggtatggta aatggtcctt tcacccagag tggcacttct caagctggct ataatgagtt
66961 tgaattggat gacttgttga atcctgatca gatgcccttc tccttcacaa gcttgctgag
67021 tggtggggat agcttattca aggttcgtca atgtgagtga tcaaatctat tttcagtttt
67081 ttttttccc tttcttccgt tcttgcagta cttagagtag aacatgaatt agaatatctt
67141 aagaaagtca tggttttgaa cagatggacc tccagcgtgt aacaagcctc tttacaattt
67201 gaattcacca attagaagag aagcagttgg gtcagtctgt gaaagttcgt ttcaatatgt
67261 accgtcaacg cccagtctgt tcagaacagg tgaaaagact ggattccttg aacagatagt
67321 tacaactact ggacatgaaa tcccagagcc gaaatctgac aaaagtATGc agagcattat
```

67381 ggactcgtct gctgttaatg cgacggaagc tactgaacaa aatgatggca gcagacaaga
67441 tgttctggag ttcgacctta acaaaactcc tcagcagaaa ccctccaaaa ggaaaaggaa
67501 gttcatgccc aaggtggtcg tggaaggcaa acctaaaaga aagccacgca aacctgcaga
67561 acttcccaaa gtggtcgtgg aaggcaaacc taaaaggaag ccacgcaaag ctgcaactca
67621 ggaaaaagtg aaatctaaag aaaccgggag tgccaaaaag aaaaatttga aagaatcagc
67681 aactaaaaag ccagccaatg ttggagatat gagcaacaaa agccctgaag tcacactcaa
67741 aagttgcaga aaagctttga attttgactt ggagaatcct ggagatgcga ggcaaggtga
67801 ctctgagtct gaaattgtcc agaacagtag tggcgcaaac tcgttttctg agatcagaga
67861 tgccattggt ggaactaatg gtagtttcct ggattcagtg tcacaaatag acaagaccaa
67921 tggattgggg gctatgaacc agccacttga agtgtcaatg ggaaaccagc cagataaact
67981 atctacagga gcgaaactgg ccagagacca acaacctgat ttattgacta gaaaccagca
68041 atgccagttc ccagtggcaa cccagaacac ccagttccca atggaaaacc aacaagcttg
68101 gcttcagatg aaaaaccaac ttattggctt tccatttggt aaccagcaac ctcgcatgac
68161 cataagaaac cagcagcctt gcttggccat gggtaatcaa caacctatgt atctgatagg
68221 aactccacgg cctgcattag taagtggaaa ccagcaacta ggaggtcccc aaggaaacaa
68281 gcggcctata tttttgaatc accagacttg tttacctgct ggaaatcagc tatatggatc
68341 acctacagac atgcatcaac ttgttatgtc aaccggaggg caacaacatg gactactgat
68401 aaaaaaccag caacctggat cattaataag aggccagcag ccttgcgtac ctttgattga
68461 ccagcaacct gcaactccaa aaggttttac tcacttgaat cagatggtag ctaccagcat
68521 gtcatcgcct gggcttcgac ctcattctca gtcacaagtt cctacaacat atctacatgt
68581 ggaatctgtt tccaggattt tgaatgggac tacaggtaca tgccagagaa gcagggctcc
68641 tgcatacgat tctttacagc aagatatcca tcaaggaaat aagtacatac tttctcatga
68701 gatatccaat ggtaatgggt gcaagaaagc gttacctcaa aactcttctc tgccaactcc
68761 aattatggct aaacttgagg aagccagggg ctcgaagaga cagtatcatc gtgcaatggg
68821 acagacggaa aagcatgatc taaacttagc tcaacagatt gctcaatcac aagatgtgga
68881 gagacataac agcagcacgt gtgtggaata tttagatgct gcaaagaaaa cgaaaatcca
68941 gaaagtagtc caagaaaatt tgcatggcat gccacctgag gttatagaaa tcgaggatga
69001 tccaactgat ggggcaagaa aaggtaaaaa tactgccagc atcagtaaag gtgcatctaa
69061 aggaaactcg tctccagtta aaaagacagc agaaaaggag aaatgtattg tcccaaaaac
69121 gcctgcaaaa aagggtcgag caggtagaaa aaaatcagta cctccgcctg ctcatgcctc
69181 agagatccag ctttggcaac ctactcctcc aaagacacct ttatcaagaa gcaagcctaa
69241 aggaaaaggg agaaagtcca tacaagattc aggaaaagca agaggtaact aatgtattct
69301 acaatctctg tgatataatt ttgagatttt agtaactgat gtgtccaaac cagctcctta
69361 tcactgttgg tgcgttgtat aggtccatca ggagaacttc tgtgtcagga ttctattgcg
69421 gaaataattt acaggatgca aaatctgtat ctaggagaca aagaaagaga acaagagcaa
69481 aatgcaatgg tcttgtacaa aggagatggt gcacttgttc cctatgagag caagaagcga
69541 aaaccaagac ccaaagttga cattgacgat gaaacaactc gcatatggaa cttactgatg
69601 gggaaaggag atgaaaaaga aggggatgaa gagaaggata aaaagaaaga gaagtggtgg
69661 gaagaagaaa gaagagtctt ccgaggaagg gctgattcct tcatcgctcg catgcacctg
69721 gtacaaggtg aagatccact tctcttctca actccatttt tattcacaca aattagtaga
69781 atactcaaaa atgatgtttt gtttgcaaaa ttttaaaatt cactagttaa ccatgtcaaa
69841 taatattcat aatgcatctt gtgaagaaca ggtgtgcatt tatggtgaca gctgaatggt
69901 ttatgtgcct attatttctt ttactgctat agatgaccaa ttgaacttaa acgtttacag
69961 gagatagacg tttttcgcca tggaagggat cggtggttga ttcggtcatt ggagttttcc
70021 ttacacagaa tgtctcggat cacctttcaa ggtatatgag ttgccttaat aaattgagtt
70081 ccaaaacata gaaattaacc catggtggtt ttacaatgca gctctgcgtt catgtctcta
70141 gctgctcgat tccctccaaa attaagcagc agccgagaag atgaaaggaa tgttagaagc
70201 gtagttgttg aagatccaga aggatgcatt ctgaacttaa atgaaattcc ttcgtggcag
70261 gaaaaggttc aacatccatc tgacatggaa gtttctgggg ttgatagtgg atcaaaagag
70321 cagctaaggg actgttcaaa ctctggaatt gaaagattta atttcttaga gaagagtatt
70381 caaaatttag aagaggaagt attatcatca caagattctt ttgatccggc gatatttcag

70441 tcgtgtggga gagttggatc ctgttcatgt tccaaatcag acgcagagtt tcctacaacc
70501 aggtgtgaaa caaaaactgt cagtggaaca tcacaatcag tgcaaactgg gagcccaaac
70561 ttgtctgatg aaatttgtct tcaagggaat gagagaccgc atctatatga aggatctggt
70621 gatgttcaga aacaagaaac tacaaatgtc gctcagaaga aacctgatct tgaaaaaaca
70681 atgaattgga aagactctgt ctgttttggt cagccaagaa atgatactaa ttggcaaaca
70741 actccttcca gcagctatga gcagtgtgcg actcgacagc cacatgtact agacatagag
70801 gattttggaa tgcaaggtga aggccttggt tattcttgga tgtccatctc accaagagtt
70861 gacagagtaa agaacaaaaa tgtaccacgc aggtttttca gacaaggtgg aagtgttcca
70921 agagaattca caggtcagat cataccatca acgcctcatg aattaccagg aatgggattg
70981 tccggttcct caagcgccgt ccaagaacac caggacgata cccaacataa tcaacaagat
71041 gagatgaata aagcatccca tttacaaaaa acatttttgg atctgctcaa ctcctctgaa
71101 gaatgcctta caagacagtc cagtaccaaa cagaacatca cggatggctg tctaccgaga
71161 gatagaactg ctgaagacgt ggttgatccg ctcagtaaca attcaagctt acagaacata
71221 ttggtcgaat caaattccag caataaagag cagacggcag ttgaatacaa ggagacaaat
71281 gccactattt tacgagagat gaaagggacg cttgctgatg ggaaaaagcc tacaagccag
71341 tgggatagtc tcagaaaaga tgtggagggg aatgaaggga gacaggaacg aaacaaaaac
71401 aatatggatt ccatagacta tgaagcaata agacgtgcta gtatcagcga gatttctgag
71461 gctatcaagg aaagagggat gaataacatg ttggccgtac gaattaaggt aaatctacta
71521 atttcagttg agaccctcat caaatctgtc agaaggcttg aacatcagta aattatgtaa
71581 ccatatttac aacattgcag gatttcctag aacggatagt taaagatcat ggtggtatcg
71641 accttgaatg gttgagagaa tctcctcctg ataaagccaa gtgggtaaat cacattttta
71701 gtgactgcaa cactagcacg atcgatttac tcaacaatta cgtcaaactg agtattaaca
71761 agttgctcat gaacatttca cagggactat ctcttgagca taagaggtct gggtttgaaa
71821 agtgttgaat gcgtgcgact cttaacactc cacaatcttg ctttccctgt gagtcagact
71881 attccattat ctactaaaaa cttagaataa ctccggctaa ctaagctgga acttgtattg
71941 atgatatgaa ggttgacacg aatgttggaa ggatagcagt taggatggga tgggtgcctc
72001 tacaacccct acctgaatca cttcagttac acctcctgga gctgtaagtt tctttttgtt
72061 tgtcatctaa acaacgaaat ttttatgcaa gtcataacca tgctgtgttt tcacagatac
72121 ccagtgctcg agtccatcca aaaatttctt tggccaagac tttgcaaact cgatcaacga
72181 acactgtatg ctcataaact ctaacaaatc atctgtctga aaaaccaata tttctttggt
72241 agaattctat tgtcattact cattactaac agcgaaatta attaacgttc tttttcttac
72301 tcaggtatga attacactac caactgatta cgtttggaaa ggtattattg ctctaagctt
72361 tgaatttatc atatggtaat ttcaagcatt gtaggcacct gatcaattat gtgtctaaat
72421 catgtgaatt catgtcaggt attttgcaca aagagtagac caaattgtaa tgcatgtcca
72481 atgagaggag agtgcagaca ctttgccagt gcttatgcta ggtaagcaag ctttcatgta
72541 cttatatgca ataattaaag ataaaattta ggattatggg taagtaacaa aaaattaggc
72601 tcagtttcat ggtagctagc tggaaatagt attacaagaa caacataaag atcaaagaca
72661 gaatcatgga tccatatgca ctatcatttt agctcttgta atccatacat gaacactata
72721 tgccaaagta gggatttcaa atatgagatt cgatgactga tgccattgta acagtgcaag
72781 acttgcttta ccggcaccag aggagaggag cttaacaagt gcaactattc cggtccctcc
72841 cgagtcctat cctcctgtag ccatcccgat gatagaacta cctcttccgt tggagaaatc
72901 cctagcaagt ggagcaccat cgaatagaga aaactgtgaa ccaataattg aagagccggc
72961 ctcgcccggg caagagtgca ctgaaataac cgagagtgat attgaagatg cttactacaa
73021 tgaggaccct gacgagatcc caacaataaa actcaacatt gaacagtttg gaatgactct
73081 acgggaacac atggaaagaa acatggagct ccaagaaggt gacatgtcca aggctttggt
73141 tgctttgcat ccaacaacta cttctattcc aactcccaaa ctaaagaaca ttagccgtct
73201 caggacagag caccaagtgt aagctaatat ctcctcctat attttatctt ccatataaat
73261 tttggggaaa aaatcgctct ccatctggtt ttagaacatg cgggtcagcc agggttatgg
73321 catttttata tatttcaccg atcggcccga gctggctctg gttgactcgt atgccaccct
73381 gcattgaaca aaccagtagg agacaagcaa gcaaaacgtt ttaagataag gtctatggta
73441 aaatgacaag gtaactgata aatgtgtcgt ctatttgcag gtacgagctc ccagattcac

73501 atcgtctcct tgatggtgta agtcaatttt taactctctc tatactcgag ttgtttcact
73561 tgagcaacac tgtttaaaag tcctcatttg ataaaataac agatggataa aagagaacca
73621 gatgatccaa gtccttatct cttagctata tggacaccag gtgagaataa aactgcaatg
73681 tttcattcat gtgtctacag tatcaaagaa agtacagcta gagctaaaaa gcatttgaaa
73741 tagagtcggt taaatatgaa agtttgaatc tgtaaatgaa agccggaacg tagcattggt
73801 ggatgttata tgtaaattag tttttgagat tggtctaatg tagttgtttg actgccaggt
73861 gaaacagcga attcggcaca accgcctgaa cagaagtgtg gagggaaagc gtctggcaaa
73921 atgtgctttg acgagacttg ttctgagtgt aacagtctga gggaagcaaa ctcacagaca
73981 gttcgaggaa ctcttctggt gagattatct tgatcttttg tgttgctcat gaaaaggaga
74041 agtgagaata caagtttgct aatatcattt tttcgtcatt cacagatacc ttgtcggact
74101 gccatgagag gaagttttcc gctcaacggg acatatttcc aagtcaacga ggttagatga
74161 ataaaactc aaacagacag acgaaacatt atttctgttt agtgttggtt ctttatcctc
74221 cttgccattt tttatcttgc agttatttgc agaccacgag tccagtctca aacccatcga
74281 tgttcctaga gattggatat gggatctccc aagaaggact gtttacttcg gaacatcagt
74341 aacatcaata ttcagaggta aaaacattcg taatagagtt agttaatcaa atgtccaaaa
74401 cacaagaaag cttcaccgtc caatacacaa gaaagcttca ccttctcttt gccaaaaaag
74461 atcttagaat gttttgctga atttgtgcag gtctttcaac ggagcagata cagttctgct
74521 tttggaaagg taaacgttaa ctttcgaccc agagaaatcc ggaaaatcta ttgctttgtt
74581 ctgatcaata cgttaaacat atacacacac actttacact taggaccaat actgttctga
74641 tctgtgatag aaactggtaa acatctaaca attatgattg caggattcgt atgtgtccgt
74701 ggattcgaac agaagacaag agcaccgcgt ccattaatgg caaggttgca ttttcctgcg
74761 agcaaattga agaacaacaa aacc**TAA**aga tgactggaag aaagcaaacg cattgcttct
74821 ctgctctcct ctatttaaag ccaggaaaag tcccatttag acataataac aggaatccaa
74881 ataggctatt ttctctttct ttcttatttc attcatagag cagaagcgac acaaaaaagt
74941 tttttgggtt atttattttc tctctaacaa atttgtagcg ttttgggtct ttttctggct
75001 gtcactagcg tggcaaatcc aatgtctgcg cacacttagg cgcattgtca ataaaatttc

SEQ ID NO:8

ARABIDOPSIS THALIANA DMT2

>DMT2(1DMT2);

MEKQRREESSFQQPPWIPQTPMKPFSPICPYTVEDQYHSSQLEERRFVGNKDMSGLDHLS
FGDLLALANTASLIFSGQTPIPTRNTEVMQKGTEEVESLSSVSNNVAEQILKTPEKPKRK
KHRPKVRREAKPKREPKPRAPRKSVVTDGQESKTPKRKYVRKKVEVSKDQDATPVESSAA
VETSTRPKRLCRRVLDFEAENGENQTNGDIREAGEMESALQEKQLDSGNQELKDCLLSAP
STPKRKRSQGKRKGVQPKKNGSNLEEVDISMAQAAKRRQGPTCCDMNLSGIQYDEQCDYQ
KMHWLYSPNLQQGGMRYDAICSKVFSGQQHNYVSAFHATCYSSTSQLSANRVLTVEERRE
GIFQGRQESELNVLSDKIDTPIKKKTTGHARFRNLSSMNKLVEVPEHLTSGYCSKPQQNN
KILVDTRVTVSKKKPTKSEKSQTKQKNLLPNLCRFPPSFTGLSPDELWKRRNSIETISEL
LRLLDINREHSETALVPYTMNSQIVLFGGGAGAIVPVTPVKKPRPRPKVDLDDETDRVWK
LLLENINSEGVDGSDEQKAKWWEEERNVFRGRADSFIARMHLVQGDRRFTPWKGSVVDSV
VGVFLTQNVSDHLSSSAFMSLASQFPVPFVPSSNFDAGTSSMPSIQITYLDSEETMSSPP
DHNHSSVTLKNTQPDEEKDYVPSNETSRSSSEIAISAHESVDKTTDSKEYVDSDRKGSSV
EVDKTDEKCRVLNLFPSEDSALTCQHSMVSDAPQNTERAGSSSEIDLEGEYRTSFMKLLQ
GVQVSLEDSNQVSPNMSPGDCSSEIKGFQSMKEPTKSSVDSSEPGCCSQQDGDVLSCQKP
TLKEKGKKVLKEEKKAFDWDCLRREAQARAGIREKTRSTMDTVDWKAIRAADVKEVAETI
KSRGMNHKLAERIQYLTLNMKIMQGFLDRLVNDHGSIDLEWLRDVPPDKAKEYLLSFNGL
GLKSVECVRLLTLHHLAFPVDTNVGRIAVRLGWVPLQPLPESLQLHLLEMYPMLESIQKY
LWPRLCKLDQKTLYELHYQMITFGKVFCTKSKPNCNACPMKGECRHFASAFARKFSNIHL
FYSARLALPSTEKGMGTPDKNPLPLHLPEPFQREQGSEVVQHSEPAKKVTCCEPIIEEPA
SPEPETAEVSIADIEEAFFEDPEEIPTIRLNMDAFTSNLKKIMEHNKELQDGNMSSALVA
LTAETASLPMPKLKNISQLRTEHRVYELPDEHPLLAQLEKREPDDPCSYLLAIWTPGETA
DSIQPSVSTCIFQANGMLCDEETCFSCNSIKETRSQIVRGTILIPCRTAMRGSFPLNGTY
FQVNEVFADHASSLNPINVPRELIWELPRRTVYFGTSVPTIFKGLSTEKIQACFWKGYVC
VRGFDRKTRGPKPLIARLHFPASKLKGQQANLA

SEQ ID NO:9

>DMT2(1DMT2) novel 480 amino acid amino terminus;

MEKQRREESSFQQPPWIPQTPMKPFSPICPYTVEDQYHSSQLEERRFVGNKDMSGLDHLS
FGDLLALANTASLIFSGQTPIPTRNTEVMQKGTEEVESLSSVSNNVAEQILKTPEKPKRK
KHRPKVRREAKPKREPKPRAPRKSVVTDGQESKTPKRKYVRKKVEVSKDQDATPVESSAA
VETSTRPKRLCRRVLDFEAENGENQTNGDIREAGEMESALQEKQLDSGNQELKDCLLSAP
STPKRKRSQGKRKGVQPKKNGSNLEEVDISMAQAAKRRQGPTCCDMNLSGIQYDEQCDYQ

KMHWLYSPNLQQGGMRYDAICSKVFSGQQHNYVSAFHATCYSSTSQLSANRVLTVEERRE
GIFQGRQESELNVLSDKIDTPIKKKTTGHARFRNLSSMNKLVEVPEHLTSGYCSKPQQNN
KILVDTRVTVSKKKPTKSEKSQTKQKNLLPNLCRFPPSFTGLSPDELWKRRNSIETISEL

SEQ ID NO:10

>DMT2 (1DMT2) Nucleotide sequence from BAC F1011 (gi 6598632);

60001 tcgctgagcc tgggtttctt catcggacct ggatctctgg atctatcaaa cggtctacga
60061 ggattctcca ttccaaagaa ctatacaata caagaggtac gcaaataatg ccctaaatta
60121 aacctaatcg gcaaaaatcg attgcagtga caacaaatcc tcgttagagg ggaattcaga
60181 gcattacaac aatcagtaac cctaagttac aatctaaaaa ttgagatgca taacgcgatt
60241 ctgcgaagaa gacggagaag atagaaggaa tgcttcgaat tcggcaaaaa tgtcagagag
60301 tttggacaat ctccgatcaa ttagggttgt gaattgggga ttttatggag acgagacaaa
60361 aaaaagttga agatcggagc tggttccaaa aatatttagg cccatttaat gacccacatt
60421 ccatgtataa taggcccatc atctaatatt tgacaacaat agaattcttt ggtccggttg
60481 aactatctga tttaaaccaa gttaagtgag atcctccaca tatcgaacca gatcttgatt
60541 caggtaacca aaagctaacc gtaaattcag atataaacca aacgaaggga acagagagtt
60601 tacacagcta cgggtctgtt ttttgtgaca agtgtttgat acaaatttaa gacgaaacta
60661 aaatgggatt tagaaacctt gtacaactct aggactgtta actttacgtt ttcactttct
60721 tacattaact agattggaac agtgtgctct ctcactctta accataagct tgtatttgtt
60781 tgcttgccaa cggaTTAggc gaggttagct tgttgtccct tcagtttgct cgccgggaag
60841 tgcaatcttg caatcaaagg cttcggtccc ctcgtctttc gatcaaatcc acgtacacat
60901 acgtacccta cataatatca aaagataagt tatgtttcag aacaagaaga aactgcttaa
60961 tacaaaatgt acctttccaa aagcaagcct gtatcttctc agttgataaa cctgagaaaa
61021 atagagctca agtggttaga acaactttct tttatataaa caatcgcatc acaatccaat
61081 aaagaaaatc ttataccttt gaatatcgta ggaacagagg taccaaaata gaccgttctt
61141 cgaggtaatt cccatatcaa ttcccttggg acattgattg ggtttaggct ggatgcatga
61201 tccgcaaaca cctgtatcaa tagaatacat cacaagtttc aatgcaaata attaaaatga
61261 aagagttgga gttattggag ttcaagtctt acctcattta cttgaaagta cgttccattt
61321 agaggaaaac taccctcat cgctgttcta caaggaatct gtacaattta caacatatta
61381 atctgtagaa aacataagtg tagtaagccg cataaggaga ttgatgcaac tacttaccaa
61441 aattgtccct ctcacaattt gagatctagt ctccttgatg ctgttgcagg agaaacaagt
61501 ctcctcgtca caaagcatac catttgcttg gaatatgcac gtactaacag acggttgaat
61561 agaatcagcc gtctcacctg ttgaataaca catcgattaa agataccgat ttgatttcat
61621 gattaaaaga tatgcaaatc attaaattac ctggcgtcca tatagcaagc aaataagaac
61681 atggatcatc aggttctctc ttttccaact gccacaagaa atcacaaaca gctagtcaga
61741 ttttacaata tagacagcac tctatacggc atgtgtcctt atccagttag ctcacatacc
61801 tgagctagaa gaggatgctc gtctggaagt tcgtaactgc aagatacggg aaaagaaaca
61861 agttatggca tagcctgtaa ttattgggaa gtttgtctgc tttccaactt acgagttcat
61921 gcttggtcaa tcacttaaat attctactct gttcaagctt taataatttt gaaaaatgtg
61981 tttctgattt cattttaac ctaagaacga agaaaaacag agaaaaatgg attcttacac
62041 tcggtgttct gtccttaact ggctgatatt cttgagctta ggcattggaa gagaagcagt
62101 ttcagcagta agtgcaacta aagcgctgga catgtttccg tcttgaagtt ccttgttgtg
62161 ttccattatc ttcttcaagt tactggtaaa tgcatccatg tttagcctga tggtaggaat
62221 ttcttctgga tcctcaaaaa acgcctcctc tatgtcagct attgatactt ctgcggtttc
62281 tggctccggt gaagcaggct cttcgatgat tggttcacaa catgtgacct tttttgctgg
62341 ttctgagtgc tgtactactt cagacccttg ctctctctgg aatggctctg gcaggtgtag
62401 aggcaaaggg tttttatcag gtgtccccat acctttctct gtacttggta aagcaagcct
62461 tgcactgtaa aacaaatgaa tgttactaaa ttttctgtaa tgatgattca gagcttcgtt
62521 tagatacaga ccaattctca tttaactggg ttatatttta acaaggactt tcctcataga

62581 gtcatagtgg tactaaaggt ttaagagaac atgttgtagc accttgcaaa cgcactggca
62641 aaatgtctgc attctccttt catcggacat gcattgcaat taggtttgct ctttgtgcaa
62701 aagacctgat acaataatca agcagattac aaacctcatc atgtgagctg attttgacat
62761 acgtatatat gtatttcttt aatacatacc tttccaaaag taatcatctg gtagtgcaac
62821 tcatacctgt gagataatag ggtattaaac taatgaataa gtgtattaga ctgaggcatg
62881 aaaaaaaaaa agttagtgat aaacatcatt cttacaatgt tttttggtcg agtttgcaga
62941 gacggggcca aagatacttt tgaatagatt caagcatagg atacctagac aaaccaaacc
63001 tcagatgtat taagtaacaa attacaattt ccaagtagga ccattttgaa aagtgcttac
63061 atttccagaa gatgcaactg aagtgactct gggagcggct gaaggggcac ccatccaagt
63121 ctgacggcta tgcgcccaac atttgtatca acctgtcaat aaattaagtt catgcatcat
63181 ataattcact ttttataggg acagaaacaa aagtttgatc cttgcttact ggaaaggcaa
63241 gatggtgaag tgttagaagc cgcacacact ccacactttt cagtcccaat ccgttaaagc
63301 tcagaagata ttctctgcag ggttttgtaa tatacgagag tacataattc attattaagt
63361 cactaaaact gccaaagtag taatctttgt ataggttaat aaagaagaaa taaaatgctt
63421 cgtctttcaa acttactttg ctttatctgg tggaacatct ctcaaccatt caagatcgat
63481 acttccatgg tcatttacca gtcgatcaag gaagccctgc atgattttca tgttcagagt
63541 caaatactta aaatgaatgt tatcacgaaa tttagccact aaatttttac ctgtatacgt
63601 tctgcaagtt tatggttcat cccgcgactc ttgattgttt cagcaacttc cttaacatct
63661 gctgctcgta ttgccttcca atccacggtg tccattgtac ttcttgtttt ttctctaatt
63721 cctgctctag cttgggcttc tcttcttaaa caatcccagt caaacgcttt tttttcctcc
63781 ttcaaaacct ttttcccttt ttcttttaag gtaggtttct gacaactcaa aacatcccca
63841 tcttgctgag agcaacaacc aggttcacta ctatcaacag aggattttgt gggctctttc
63901 attgactgga aacccttaat ttctgagcta caatcacccg gagacatatt tggtgatact
63961 tgattggaat cttctagaga gacttgtacc ccctgtagga gcttcataaa ggaagtacga
64021 tactctcctt ctaagtcgat ctctgagctt gatcctgctc tctctgtatt ttgaggagca
64081 tcagacacca tcgaatgttg acatgtaagt gcagaatctt cagatggaaa caggttcagg
64141 acacgacact tctcatccgt cttatcaacc tctacacttg agccttttcg atctgaatca
64201 acatactcct ttgaatccgt ggttttgtca actgattcat gggctgagat ggcaatctca
64261 ctactgcttc tggaggtttc attgctaggt acataatcct tctcctcatc aggctgtgta
64321 tttttcaaag taacagaact gtgattgtga tcgggtgggc ttgacatcgt ttcctctgag
64381 tccaagtacg ttatttgaat agaaggcatc gagcttgttc cagcgtcaaa gttactgctc
64441 ggtacaaaag ggacagggaa ctgggaagcc aacgacatga aagccgaact acaaggagta
64501 aaaaacatca aagcaagtta gttttgtgac tttttgctgt cttggattta gtttgacata
64561 gaattatgta agagcttgta ccttgagaga tggtctgaaa cattttgagt gagaaatact
64621 ccaacaacag aatccacgac ggatcccttc caaggcgtaa aacgtcgatc ccctgttaga
64681 aaccaaagac cataacaaga agcagtagct gagacatact aattgaaacc atgtggttag
64741 aacagaaaca cataaaagga caagtgtggt gtataacctt gtacaaggtg catccttgca
64801 ataaatgagt cagctcgtcc tcgaaacaca ttacgttctt cctcccacca tttcgccttc
64861 tgctcgtctg atccgtcaac accttcgcta ttaatattct ccaatagcag tttccacact
64921 ctgtctgtct catcgtctag atcaaccttt ggtcgtgggc gtggtttttt aacaggagtt
64981 acaggcacaa ttgctccagc gccaccacca aagagtacaa tctggctatt cattgtgtaa
65041 ggaacgagag cagtttcaga atgctccctg ttgatgtcta atagacgcaa tagctcactg
65101 attgtttcga tcgagttacg tcgtttccaa agttcatctg gagaaagacc tgcaggaatc
65161 aaacatcatc attatcaaga aatagtctgc atttaacaga ttcaaaaaaa caaagaaata
65221 tagttctgta tctattcatt accagtaaat gaaggtggaa aacggcaaag attcggaaga
65281 agatttttct gtttggtttg tgatttctca gacttggttg gcttcttttt gctcacagtc
65341 acccgcgtat caacaagaat cttattattt tgctgtggct tgctacaata tcctgaggtt
65401 aaatgctcag gaacttccac aagtttattc attgaagaca aattccggaa tcgagcatgg
65461 cctgttgttt tcttcttgat cggcgtgtct atcttatccg agagaacatt tagctcagac
65521 tcttgccttc cttgaaagat accttctcgt ctttcttcaa cggttaggac tctattagca
65581 ctgagctgag atgtggaact gtagcacgta gcgtgaaagg cagaaacata attgtgctgt

65641 tgtccagaga atactttgct gcaaatggca tcatatctca tccctccctg ttgcaagttt
65701 ggggaataca accaatgcat tttctggtag tcacattgct catcatactg aatccctgat
65761 agattcatgt cgcaacaagt tggtccttgt cttctctttg cagcttgcgc catcgaaata
65821 tcgacttctt ctagattact gccatttttc tttggttgaa ctccctttct tttaccttgg
65881 ctgcgctttc tcttgggcgt gctaggagcc gaaagaaggc aatcttttaa ctcttgattc
65941 ccagaatcta actgcttctc ttgaagagct gattccatct cacctgcttc tctaatgtca
66001 ccgttggtct ggttttctcc attttcggct tcaaaatcca agactcgtct acagagcctc
66061 ttaggacgag ttgaagtttc aacagctgct gatgattcaa ccggagtagc gtcttgatcc
66121 ttactgactt caaccttctt ccgcacatat ttcctctttg gtgttttgct ttcttgacca
66181 tcggtgacaa cagacttcct cggagctcgt ggtttaggct ccctcttggg tttagcttct
66241 ctacgaacct ttggccgatg cttcttcctc ttaggttttt caggagtctt gaggatctgt
66301 tcagcaacat tgttactcac tgagctcaaa ctctccactt cttcagtacc tttttgcata
66361 acctctgtgt ttcctacatt gagaatcaca tctttctcag tccaactcaa acagaatcaa
66421 aatttgacaa agcgatttca tttctcatga gaccagaatc aaaatcccct cttacttgta
66481 ggtattggag tctgaccaga gaatatgagg gatgcagtgt tagctagagc aagcaaatcc
66541 ccaaaagaca agtgatcaag accactcata tccttgttcc caacaaatct cctgcatgca
66601 tcaatacctt acttaaccaa ttacccatca ctactctttg aaatttctca actttagaac
66661 aaaaaagcac aaacctttcc tccaattgac tgctatgata ttgatcctcc accgtgtatg
66721 ggcagatcgg tgaaaatggc ttcatgggtg tctgaggaat ccatggaggt tgttgaaagc
66781 tgctttcttc tctcctctgt ttctcCATtt ctgactctat ttttactttt cttcactctt
66841 acttaaatca gaaccatttg agaaaaagct tggaactttc tattttttcc actgcaaaaa
66901 gttcaataat ttcttcaata aaagagatca ccaatttttt ttaaaaatca cgattttata
66961 aaatgatcag atccactttt ttctggggtt ttagagaaag agagatctcc ggaagtcatt
67021 gattttgggt gagtggcgac atgaacgatt aatccgttcg ttaggtgaaa gagagacttt
67081 ttagattcac aacaaaatgt aaaaaaaagt aagaaaaaaa caaaattcat taccagtaga
67141 atcaatggtt atggtggtga tggagagagt tagttcggtg gtagctatga gaggataaga
67201 tcactgatgc ttcgtttctt ctcttggaat cgatgaagtt aaagagtaat atagaaaaag
67261 ctttttggc ctaacgtata aagaagagga tataacatgt gttgttgtgt gtttcactat
67321 ttttcataac cgtttgttta tgtagggcga aagttcgttt ggttggcggg aaaagtttta
67381 cggaatttta ttttaaaaat aatgattctt ttctacaaaa tctcctagac tatgggaaag
67441 atgatttaaa aagttaataa tattgtcgtt gttatcgtca tcgtcatcat cgtcttttct
67501 gttatctttt tctctttaaa atttcgtatt ttttctcgtt tacgtaacta tttaaaatta
67561 tatgaactaa ctaattttat aattaataga aattataaaa taatcttaat tttgctttag
67621 atataaaata attagaactt tatttataaa tttatcatca aattatgatt taaacaaata
67681 acatgttatg taatccacgt ttataatttt gatcaataat atattatttt gctaattttt
67741 acgtaatctc ataaatttac acgttttcgt ttacatatgc agaagttaaa tgattcgttt
67801 tagaattatt attttccact gatatgggag ctagtgtagt agagtgatta ttaggctagt
67861 tgcccaacga gtctttcgtt tttgatcatt ccaaatgttt tagtctagta cgataggagt
67921 caaaatactg caccatatgt gtgaaactgt gaatgtgtgt gaaaaaaaga gtaattagtg
67981 tgctaacctt tgatttcctg tcatgcaaga aaccttcaaa gagacgtaca tgagaaatga
68041 gtattgtaaa tcatttattt catggacttg gttggaatct tagtgaatcg ttgttgtcaa
68101 tcttaacaac ttgttggatt ggttatgagc ctatgactta tgacttatga gtgagtcaat
68161 ggtggtcata acctaatgat tgggttatga gcaaagaaat ttggaatttg taaaaaaaaa
68221 aaaaaaaatc aagagctttt ttgtgtggac atatctatcc tagaaactga gacgaataat
68281 agtggataaa aagttgggaa cggattattc gaatgtttaa aactattatt gaaaacaata
68341 caactaaata tggtacaaaa gtaaacgaat tcgtatagct aaacctaatt caaattacga
68401 agctaatcca tacttggatc ctaaacgctt ttacttttac ttacggtttc tttttcaaaa
68461 aagtttttac aaatttgggt ttgtcttatg aagattatgg cagaagagac tgatcaaaag
68521 tgaatgccta attcggttta atccattcaa gtttatctta aacaatgaaa ctgaccatga
68581 aagtgaattc aaagaccaaa tcaaagaaaa attaaactga tttagttgta atattggtat
68641 tgaattaaac tataaataga aataaccaaa catataacca caaaagaaga ctatttatat

68701 aaatatatga gttggaagtc atttttggac tattatataa gatctaatta tcacacgacg
68761 tgtggatgta tggttagcag agttgtgttc agagagttcg ataaagccat cactccaaac
68821 atacaaaata tccatacatt gatccaccaa tataaccggc tgtgtgccaa gcaaagtgaa

SEQ ID NO:11

ARABIDOPSIS THALIANA DMT3

>DMT3(1DMT3);

MEVEGEVREKEARVKGRQPETEVLHGLPQEQSIFNNMQHNHQPDSDRRRLSLENLPGLYN
MSCTQLLALANATVATGSSIGASSSSLSSQHPTDSWINSWKMDSNPWTLSKMQKQQYDVS
TPQKFLCDLNLTPEELVSTSTQRTEPESPQITLKTPGKSLSETDHEPHDRIKKSVLGTGS
PAAVKKRKIARNDEKSQLETPTLKRKKIRPKVVREGKTKKASSKAGIKKSSIAATATKTS
EESNYVRPKRLTRRSIRFDFDLQEEDEEFCGIDFTSAGHVEGSSGEENLTDTTLGMFGHV
PKGRRGQRRSNGFKKTDNDCLSSMLSLVNTGPGSFMESEEDRPSDSQISLGRQRSIMATR
PRNFRSLKKLLQRIIPSKRDRKGCKLPRGLPKLTVASKLQLKVFRKKRSQRNRVASQFNA
RILDLQWRRQNPTGTSLADIWERSLTIDAITKLFEELDINKEGLCLPHNRETALILYKKS
YEEQKAIVKYSKKQKPKVQLDPETSRVWKLLMSSIDCDGVDGSDEEKRKWWEEERNMFHG
RANSFIARMRVVQGNRTFSPWKGSVVDSVVGVFLTQNVADHSSSSAYMDLAAEFPVEWNF
NKGSCHEEWGSSVTQETILNLDPRTGVSTPRIRNPTRVIIEEIDDDENDIDAVCSQESSK
TSDSSITSADQSKTMLLDPFNTVLMNEQVDSQMVKGKGHIPYTDDLNDLSQGISMVSSAS
THCELNLNEVPPEVELCSHQQDPESTIQTQDQQESTRTEDVKKNRKKPTTSKPKKKSKES
AKSTQKKSVDWDSLRKEAESGGRKRERTERTMDTVDWDALRCTDVHKIANIIIKRGMNNM
LAERIKAFLNRLVKKHGSIDLEWLRDVPPDKAKEYLLSINGLGLKSVECVRLLSLHQIAF
PVDTNVGRIAVRLGWVPLQPLPDELQMHLLELYPVLESVQKYLWPRLCKLDQKTLYELHY
HMITFGKVFCTKVKPNCNACPMKAECRHYSSARASARLALPEPEESDRTSVMIHERRSKR
KPVVVNFRPSLFLYQEKEQEAQRSQNCEPIIEEPASPEPEYIEHDIEDYPRDKNNVGTSE
DPWENKDVIPTIILNKEAGTSHDLVVNKEAGTSHDLVVLSTYAAAIPRRKLKIKEKLRTE
HHVFELPDHHSILEGFERREAEDIVPYLLAIWTPGETVNSIQPPKQRCALFESNNTLCNE
NKCFQCNKTREEESQTVRGTILIPCRTAMRGGFPLNGTYFQTNEVFADHDSSINPIDVPT
ELIWDLKRRVAYLGSSVSSICKGLSVEAIKYNFQEGYVCVRGFDRENRKPKSLVKRLHCS
HVAIRTKEKTEE

SEQ ID NO:12

>DMT3(1DMT3) novel 375 amino acid amino terminus;

MEVEGEVREKEARVKGRQPETEVLHGLPQEQSIFNNMQHNHQPDSDRRRLSLENLPGLYN
MSCTQLLALANATVATGSSIGASSSSLSSQHPTDSWINSWKMDSNPWTLSKMQKQQYDVS
TPQKFLCDLNLTPEELVSTSTQRTEPESPQITLKTPGKSLSETDHEPHDRIKKSVLGTGS

PAAVKKRKIARNDEKSQLETPTLKRKKIRPKVVREGKTKKASSKAGIKKSSIAATATKTS
EESNYVRPKRLTRRSIRFDFDLQEEDEEFCGIDFTSAGHVEGSSGEENLTDTTLGMFGHV
PKGRRGQRRSNGFKKTDNDCLSSMLSLVNTGPGSFMESEEDRPSDSQISLGRQRSIMATR
PRNFRSLKKLLQRII

SEQ ID NO:13

>DMT3(1DMT3) nucleotide sequence from BAC T22K18 (gi 12408726);

53341 aatcaagtac taatgcagat ttaaggggg tgtattgacg gcgttaaaac ggtttctcaa
53401 cggaatcgta cgtagtcaca cgtgatttta ttgtttaccc cggattggtc atgcgttcct
53461 tcttttccac ttgcgcggac cactcaatga cactctcttc ttttgtagca gtggcccgac
53521 accagaatgc agcatttaat ctctcaaatt accattttgc tcctacctct tttacccctt
53581 ttggtatttt gtgtcttttt tctttctatt tcgtgtgaaa aaggatctct tccttaatcg
53641 tattatttct tccgatatct acttttattc tgttttctat ttttggtagg ttacatcttt
53701 tttataaaga aaatatgagc taacacgaca ttagtgttgt taaccaaaga attggaaaaa
53761 agttataaga gagataataa gattctctta cagagactca cttcagtgaa aaaggaagaa
53821 gcaagtggtt cccttaaggg aaaaaaaagt cacgtacgtt catatacaac tttaatacgt
53881 actgtgtaac tcaatagatc gtgcagtaat attcagtcgt attagtaaga aggaatttat
53941 ttgctaagta aactcaagcc tccttttttct ctttttttttc tttttagtaa aaattaggct
54001 agtgtttttt ttgactcagc aacactctgc ttaaatttag gagtaatttg acctattcct
54061 acgagtttct aagtgaattc tgttggggtc aaagaagcaa ctagttgaat tagtggaaaa
54121 tcgtttcctt tctttacgca tagttcacgt tggacactca gtctcaatgc tttcacgttt
54181 cacgtagcaa caacatatat tcatcagttt gtgatcgtgc catcgtggat aagttgcaat
54241 tcagtgaaac tctgcaccac tttgtgcaat tatttggccg tctaatctat ttgtgagaat
54301 tttacaatct aattgttcta ttatttcatt tacttgtcat caatttatta tatttgtagc
54361 caatgaacgt tgtaattaaa gaaccaaaat aaattaatat cttgaaattt gtaacagtca
54421 ctagaagctg atttcttatt aattgtatca ctaaagtatt attaaaaacg gttacaaatt
54481 atgataatta tatatttaat aaatttcgtg tgtcacattt cttttaaact acaattatga
54541 atatctaaaa ctcattcatg catatcttaa aatttgaatt caaaactttc ttatcttatc
54601 tttaggttct taattaacag tcactaaaaa tagtcaaagt tttgaagttt atgaaaaaag
54661 ataagagtat aattaatgga tacgcctcgt aacaaattct tgtaaagtat agataatata
54721 catttgttaa atatgacacg tgtttatttt ttttttaaat atgatcaaaa tatattttaa
54781 ctacctagat ggtatgtatg tctccaattt tgaataacaa gtcaattgtt attagaaatg
54841 tcataatata aagaagggaa ttaaatttgc aaagaaaaag tgaaaaacaa aggatttgta
54901 ttttggagaa aattaaggac tggatttgca aaaacgaaaa agtaacttca tgtatattgt
54961 cttccttata gtctctataa actattatct caaattttgt ctggactctg aaactcacaa
55021 gacttgactc tggcttactt ggcttcatct ttttctctct ggtaatctct cctgcaactt
55081 caagctttca ttttcaaata aatgtaatca aatctgttat tttcactcaa gaactaattg
55141 agttctctat ccctttcaat tgaaattgac attaaaatga aaagattttg aggaggtttc
55201 acctaccaca accgaatcac ttctttctcc aaatattgtt tctttcagtg gccaagaatc
55261 acaatcaatt tttgtatctt ccacaggtaa attaattgtg attgaacaga gaagaggaca
55321 agtgatcttg gttcaaaaga aATGgaagtg gaaggtgaag tgagagagaa agaagctagg
55381 gttaaaggga gacaaccaga gacagaagtt ctacatggtc tgccacaaga acagtcaata
55441 tttaataaca tgcaacacaa ccatcagcct gactcagaca ggttttgtga ctcaaccgaa
55501 tttactctgt tcttctcccg gaatttccat attttctggt gattctgttt tgttaaattc
55561 tgcaaaagga agaaaataaa tcaaacattt ttcacttctt caaaacatga gtaaatgcaa
55621 aaactgagat atgtaaacac acagcaattt tttgatgaac tggttttggc tgtgtgatct
55681 ttgtgtctat gcaattacgt tttagttatt ttctacttta taaggagaga tgttaactga
55741 aactgttatt gatcatacag gaggaggctt agtcttgaaa acttacctgg actatacaac
55801 atgtcttgta cacaactctt ggctctggcc aatgccacag tcgccacagg ttcatcaatt

55861 ggtgcatcat catcatcgtt aagctctcag catccaacgg attcttggat taatagctgg
55921 aagatggact ctaatccgtg gactttgagt aaaatgcaaa aacaacaatg tgagtaaaat
55981 ttgttcctga atttgtagga tcttttaaga gaaagtaagc gtttatgtgt agattaagtc
56041 agactgaaat cgattatctc ataataagtt ctcagtgatc tctcaaatca tgaattttat
56101 gtttacctga tatcaacttc ttgtcttggt gaaccacaga tgatgtgtca actccgcaga
56161 agtttctttg tgaccttaat cttacacctg aagagttggt gagcaccagt acgcaacgaa
56221 cagaacctga gtctcctcaa ataactttaa agacaccagg aaaaagtctg tctgaaactg
56281 atcatgagcc tcacgaccgt atcaagaagt ctgttcttgg aactggatct cctgcagcag
56341 taaagaaaag aaagatagca agaaatgatg agaaatctca gctggaaaca ccaacactaa
56401 agagaaaaaa gatcaggcca aaggttgtcc gtgaaggcaa aacaaaaaaa gcatcatcta
56461 aagcagggat taaaaaatcc tctattgctg ctactgctac taaaacttct gaagagagca
56521 attatgttcg gccaaaaaga ttaacgagaa gatctatacg attcgacttt gaccttcaag
56581 aagaagatga ggaatttgt ggaatcgatt tcacatcagc aggtcacgta gagggttctt
56641 caggtgaaga aaatctaacc gatacaacac tgggaatgtt tggtcacgtc ccaaagggaa
56701 gaagagggca aagaagatcc aatggcttta aaaaaaccga caatgattgc ctcagttcta
56761 tgttgtctct tgtcaatacc ggaccaggaa gtttcatgga atcagaagaa gatcgtccga
56821 gtgattcaca aatttctctg ggaagacaga gatccattat ggcaaccaga ccgcgtaact
56881 tccgatcgtt aaagaaactt ttacaaagga ttataccaag caaacgtgat agaaaaggat
56941 gtaagcttcc tcgtggactt ccgaagctta ccgtcgcatc caagttgcaa ctaaaagtgt
57001 ttagaaagaa gcggagtcaa agaaaccgtg tagcaagcca gttcaatgca aggatattgg
57061 acttgcagtg gcgacgccaa aatccaacag gtgataaaca cacaagcaac tttcatctat
57121 aatatttttc ttagatttct atcttttgaa ttaatactag ttttacaaaa tgcaggtaca
57181 tcgctagctg atatatggga aagaagtttg actattgatg ctatcactaa gttgtttgaa
57241 gaattagaca tcaacaaaga gggtctttgc cttccacata atagagaaac tgcacttatt
57301 ctatacaaaa agtcgtatga agagcaaaag gcaatagtga agtatagcaa gaagcagaaa
57361 ccgaaagtac aattggatcc tgaaacgagt cgagtgtgga aactcttaat gtcaagtatc
57421 gactgtgacg gtgttgatgg atcagatgag gaaaaacgta aatggtggga agaggagagg
57481 aacatgttcc atggacgtgc aaactcgttc attgcgcgaa tgcgtgttgt ccaaggtatt
57541 atttattgct ttagttatga cattgttgtg tggctttata ccttagatct ttctttcttt
57601 ctttttgta tccaaagcaa catggtctta aatcaagctt atcactgcag gcaatagaac
57661 tttctcacct tggaaagggt cagtagtgga ttcagtagtg ggagttttcc taacccagaa
57721 tgtcgcagac cattcatcaa ggtatatgca ttcaagagat ttctaataag tagaagatat
57781 atgcaacaga gtggtttaga aattataact tgttcacttt tgcagttctg catatatgga
57841 tttagctgct gagtttcctg tcgagtggaa cttcaacaag ggatcatgtc atgaagagtg
57901 gggaagttca gtaactcaag aaacaatact gaatttggat ccaagaactg gagtttcaac
57961 tccaagaatt cgcaatccaa ctcgcgtcat catagaggag attgatgatg atgagaacga
58021 cattgatgct gtttgtagtc aggaatcctc taaaacaagt gacagttcca taacttctgc
58081 agaccaatca aaaacgatgc tgctggatcc atttaacaca gttttgatga acgagcaagt
58141 tgattcccaa atggtaaaag gcaaaggtca tataccatac acggatgatc ttaatgactt
58201 gtcccagggg atttcgatgg tctcatctgc ttctactcat tgtgagttga acctaaatga
58261 agtaccacct gaagtagagt tgtgcagcca tcaacaagac ccggagagta ccattcagac
58321 acaagaccag caagagagca caagaacgga ggatgtgaag aagaatagga aaaaaccaac
58381 tacctccaaa ccaaagaaaa agtcaaagga atcagcaaag agcacgcaaa agaaaagcgt
58441 tgactgggat agtttgagaa aggaagcaga aagtggtggc cgaaagagag agagaacaga
58501 aagaacaatg gacacagttg attgggatgc acttcgatgt acagacgtac acaagatcgc
58561 taatataatc atcaaacgag ggatgaacaa catgcttgcc gaaagaatca aggtttgact
58621 aatcacagtg ctatatatac ctcatttata cattctaaca aggtgaattt ttttgactct
58681 ggaaattgga caggccttct taaacagact agttaaaaaa catggaagca ttgacttaga
58741 gtggctaaga gatgttcctc ctgataaagc caagtaagaa aattatttac aaatcttgag
58801 attatatgta gcctctggtt aaagaatata tctcagtaaa tggaatcgat agtaattgag
58861 atacatataa atgagagata cttgatagtg actactaatg gttgcaggga gtatctacta

```
58921 agcataaacg gattaggatt gaagagtgtg gagtgtgtta gacttttgtc actacatcag
58981 attgcattcc ctgtaagtca atgaaggata ctgaatactc agaccctaat gaatgtggaa
59041 cagatacatt aatagttacg tatttttaca aatgcaggtt gacacgaatg tcggacgcat
59101 agctgtaaga ctaggatggg ttcccttaca gccattgccc gacgagctgc aaatgcatct
59161 tttagagttg taagaaaaaa aaattaaaga tcattcttca atcatgaaag ggaacatgag
59221 aaatttacag tagttccctt taattctatt caggtaccca gttctagagt cagttcaaaa
59281 gtacctctgg ccacgcctct gcaagcttga ccaaaaaacc ttgtaagtaa attacattag
59341 catcaaccat tactctagac ccttaaactt ctctaactaa ctctaactgt atcatacaat
59401 tctaggtacg agctgcatta ccacatgata acatttggaa aggtacctca aacaaatttc
59461 aagtgtttgt ggaatgaaaa catcttaaag tggcttttcc tattttgcag gtcttttgca
59521 caaaagtaaa acccaattgc aatgcatgtc caatgaaggc ggagtgtcga cattactcta
59581 gtgcacgtgc aaggttaaac cccacaaaat tctttgttat tgccattaac atgaaaaaaa
59641 aaacactagc ttaaagagaa agagatctgc tcaaaatagt cattttaatg gttgtatgtt
59701 ctaaatgctt gtgttatatc gcagcgcacg gcttgcttta ccagaaccag aggagagtga
59761 cagaacaagt gtaatgatcc atgagaggag atctaaacgc aagcctgttg tggttaattt
59821 tcgaccatcc ttatttcttt atcaagaaaa agagcaagaa gcacaaagat cccaaaactg
59881 tgaaccaatc attgaggaac cagcatcacc agaaccagag tatatagaac atgatattga
59941 agactatcct cgggacaaaa acaacgttgg aacatcagag gatccttggg aaaataagga
60001 cgtaattcct accatcatcc tcaacaagga agctggtaca tcacatgatt tggtggtcaa
60061 caaggaagct ggtacgtcac atgatttggt ggtactaagc acatatgcag cagcaatacc
60121 tagacgtaaa ctcaagatca aggaaaagct acgcacagag caccacgtgt gagttgccac
60181 tttcaatttt ttcttctatt ataccctaaa ccgtaaaatt tgagactttc ctcagcattt
60241 atctcatact aattctcttt tacagatttg agctccctga tcaccattcc attctagaag
60301 gggttagtaa ctcttgcaaa atgatttagc aagaattttt ctacttattc ccgccttaaa
60361 aactgtttga ttatcttttt ttacagtttg agaggcgaga agctgaggat atagtccctt
60421 acttgttagc catttggacg ccaggtaaga agaaataggc acacaataaa atctgattat
60481 gatttttctt ttcaagaata ccgctatatt tttacgagtt ttcatcctta gatgtatatg
60541 actaatgtct aacaagtgat tgtaatattt ttccatacca ggtgaaaccg tgaattccat
60601 tcaaccgcca aaacaaagat gtgctttatt tgaaagcaat aatacattat gcaacgaaaa
60661 caaatgtttt caatgcaaca agacacggga agaggaatca cagactgtac gaggaactat
60721 attggtaaga ttctggtgga caattttcaa gagaatatct ctaagtagaa atataaggaa
60781 ggtataaaaa tgactaattt gtttgttaac agataccttg cagaacagca atgagaggtg
60841 gattcccttt gaatggcaca tacttccaaa ctaatgaggt aattttccca aaaatgaatt
60901 taacttaaac aaatgatcaa aagcaacatt ctcgtcaaag ctcgatttgg actatacttg
60961 tgcaggtttt tgctgaccat gactctagca taaaccctat cgacgtccca acagaactga
61021 tatgggatct aaaaagaaga gtcgcatact taggatcctc tgtatcctcg atttgtaaag
61081 gtaaattttc aaaacaaaac tgtcgattta tgcatgtgtt tggatatata aatccaaggt
61141 cttgtctcaa tatgtttttc tcatttttt aggtttatca gtggaagcca taaaatacaa
61201 tttccaggaa ggtatgctaa tatgtcttac actgaaaaca cctttagtat caaacattga
61261 attcatgaaa agaacaaaca atagtatcaa aatcagtcac gatgtttttg ctttggcgat
61321 gtaagatgtt gataggaaag tatagaagat atagcttaag ttggttaata ctgtttttat
61381 agagctttga ggtggggttt gactagcatt gtaatatata tgcaggatat gtctgtgtaa
61441 ggggattcga cagggagaat cgtaagccaa agagtctagt gaaaagactg cattgttctc
61501 acgtagcaat cagaactaaa gagaagacag aggaatgaaa ccttccagat tgcattaaca
61561 tgttagacat atttgattca ttggtttagg gtttacatca ccaaggtcat agaggatctt
61621 agcttttcat taacttttaa attcatgcaa ctctttttag gtgtttcttt ttgttccttg
61681 ccatagtttt gggcaatgga tggatgttct ttgcaaactc aggtttttg tagtcattaa
61741 cagaaatttg cagcactaat tcatctttcc tattatctat caaagctctc agtgtttctc
61801 cataacttga tgagatttag tcactctcaa gctaattcag tctggtccta atttcaatca
61861 gatttggtaa aggaacaact gcaattgcta agtacaaatc gatccagatt tcaaacaagt
61921 tccaggttta atccaaatca tcacattcaa tcaaagacca aactagaatt caaaacatat
```

61981 aatctctgat tcagattcaa gaaagacaaa gcatgagaca tcattctgca agttaaccaa
62041 ttccggttat tctcgaatcc tactgaatta agcatcaatc atctaaagga acttcataag

SEQ ID NO:14

Arabidopsis thaliana DMT4

>DMT4(1DMT4);

MEFSIDRDKNLLMVVPETRIKTKQFEKVYVRRKSIKLPQNSMVHNTLIKMARQRIQKSMK
ESVMNQHIFKNFDSYLSVIYHPCCFVINNSQTTHKKKEKKNSKEKHGIKHSESEHLQDDI
SQRVTGKGRRRNSKGTPKKLRFNRPRILEDGKKPRNPATTRLRTISNKRRKKDIDSEDEV
IPELATPTKESFPKRRKNEKIKRSVARTLNFKQEIVLSCLEFDKICGPIFPRGKKRTTTR
RRYDFLCFLLPMPVWKKQSRRSKRRKNMVRWARIASSSKLLEETLPLIVSHPTINGQADA
SLHIDDTLVRHVVSKQTKKSANNVIEHLNRQITYQKDHGLSSLADVPLHIEDTLIKSASS
VLSERPIKKTKDIAKLIKDMGRLKINKKVTTMIKADKKLVTAKVNLDPETIKEWDVLMVN
DSPSRSYDDKETEAKWKKEREIFQTRIDLFINRMHRLQGNRKFKQWKGSVVDSVVGVFLT
QNTTDYLSSNAFMSVAAKFPVDAREGLSYYIEEPQDAKSSECIILSDESISKVEDHENTA
KRKNEKTGIIEDEIVDWNNLRRMYTKEGSRPEMHMDSVNWSDVRLSGQNVLETTIKKRGQ
FRILSERILKFLNDEVNQNGNIDLEWLRNAPSHLVKRYLLEIEGIGLKSAECVRLLGLKH
HAFPVDTNVGRIAVRLGLVPLEPLPNGVQMHQLFEYPSMDSIQKYLWPRLCKLPQETLYE
LHYQMITFGKVFCTKTIPNCNACPMKSECKYFASAYVSSKVLLESPEEKMHEPNTFMNAH
SQDVAVDMTSNINLVEECVSSGCSDQAICYKPLVEFPSSPRAEIPESTDIEDVPFMNLYQ
SYASVPKIDFDLDALKKSVEDALVISGRMSSSDEEISKALVIPTPENACIPIKPPRKMKY
YNRLRTEHVVYVLPDNHELLHDFERRKLDDPSPYLLAIWQPGETSSSFVPPKKKCSSDGS
KLCKIKNCSYCWTIREQNSNIFRGTILVFADHETSLNPIVFRRELCKGLEKRALYCGSTV
TSIFKLLDTRRIELCFWTGFLCLRAFDRKQRDPKELVRRLHTPPDERGPNGFHIVVVDEK
EESPRVGLMVMPGFWIGGSVIQNRVYVSGVKVLE

SEQ ID NO:15

>DMT4 novel 372 amino acid NH2 terminus;

MEFSIDRDKNLLMVVPETRIKTKQFEKVYVRRKSIKLPQNSMVHNTLIKMARQRIQKSMK
ESVMNQHIFKNFDSYLSVIYHPCCFVINNSQTTHKKKEKKNSKEKHGIKHSESEHLQDDI
SQRVTGKGRRRNSKGTPKKLRFNRPRILEDGKKPRNPATTRLRTISNKRRKKDIDSEDEV
IPELATPTKESFPKRRKNEKIKRSVARTLNFKQEIVLSCLEFDKICGPIFPRGKKRTTTR
RRYDFLCFLLPMPVWKKQSRRSKRRKNMVRWARIASSSKLLEETLPLIVSHPTINGQADA

SLHIDDTLVRHVVSKQTKKSANNVIEHLNRQITYQKDHGLSSLADVPLHIEDTLIKSASS
VLSERPIKKTKD

SEQ ID NO:16

>DMT4 nucleotide sequence BAC F28A23 (gi 7228244);

```
14881 gctatggatg tcaacagaga gaattacgaa ttgggtttac cgatcattga gaaagccggc
14941 gttgctcaca agatcgactt cagggaaggc cctgctcttc ccgttcttga tgaaatcgtt
15001 gctgacgtaa gcattcttct ttctgacgta attaacaaaa aagatgatga agataatgaa
15061 ataattaaaa actcatggcc taattaggtt gatttaatat cttgatgaga atttctgtat
15121 acgcaaattt gtttcctttt tcatagaaga aagtgtggta actgattatt gtgtgtggtt
15181 gggtgcagga gaagaaccat ggaacatatg actttatatt cgttgatgct gacaaagaca
15241 actacatcaa ctaccacaag cgtttgatcg atcttgtgaa aattggagga gtgattggct
15301 acgacaacac tctgtggaat ggttctgtcg tggctcctcc tgatgcacca atgaggaagt
15361 acgttcgtta ctacagagac tttgttcttg agcttaacaa ggctcttgct gctgaccctc
15421 ggatcgagat ctgtatgctc cctgttggtg atggaatcac tatctgccgt cggatcagtt
15481 gatttgactc ctccctactc tgagtttgtc cacagtggat tactttccat cttcttatac
15541 ctttcaatcg cattttcacc aaccactaaa atggaccttt ttatgtattt gtgttaagta
15601 atatctccat tgtccttgtt ttgctttctt ctgaacaaag aaataatatg taccttactt
15661 ttcttcttgg tctcgttctt ttgtttttct ccatgataca acatctaaag aaattatttg
15721 tgtcacagca acgtaagtcg ataaaattag ttgaacatat tgagaaaaag ttatcataga
15781 ccttcaattg ttgaaagtcg atgttggtat ttgtcaattg atattagatt accaaataaa
15841 tattagacag taagaaacga acaaagtagg aagatgtagg tcaccggtct ttgaaaattt
15901 atcagataga attcataata cacagttagg tagtttcagt tgagagttaa aagggaaaaa
15961 tatgtaattg tgtgtgataa atacgtcaaa aattagttga tgagcaaaat cgtaaacaaa
16021 aatacttttt tgcattagtt ttgttggatt ccctataaat acgggttccc atatctaact
16081 cgtagttagc ataattataa gcaacaaata aacacaaaat actgaattta gaaattttcc
16141 agaaaattaa ttagagattt tacattattt ttacaaactt tagtgaatta tttcttaaac
16201 gtatgttagt tatttattaa ctgaagtttc acatatttga tagaataaca tttaaataaa
16261 aaaatttgaa gtaaggttag aatgttctta taatacttta taacttttttt aaaaggtaca
16321 agccaaaatt atcgcaaatg taaataataa atcattgtaa aaatcttaaa ctaattaaaa
16381 gatctaacgc aatctaaaca aagatttggt atcatcgccc atttatgttt tgatataatc
16441 aaaactggtt aataattaaa ttaaattatc aatttcttaa ttagttagaa ttcttgttaa
16501 tgtaatcaac tcaccattat tttaattatt taaaatatgg gttaatatct cttaatcata
16561 tctaagatga tattttcttc catttatgaa aagaaaaata tgttaattaa gcattaaaaa
16621 gaaggaaaaa ataatttaaa taatattaaa tatatataca tcgtttttag agttcgagtt
16681 cttccgtatt tacagtttct ctttttttcca aagcagggtt tggattggta gtttttctgg
16741 attaattttg tctcaaattc tttcttcttt ttatttttttt ttgtgaaatt ctttgtttta
16801 attggtgtga catcgtttcc aaaatatttt caaatttgat tgcttttgaa gttttttttt
16861 tttttctatg ttttggaatt cattatacta gcgttgttgt ttttctttct gcaagagtaA
16921 TGgagttttc aatagatcga gacaaaaatc ttctcatggt tgttccagag acacgtatca
16981 aaacaaaaca atttgaaaaa gtttatgtga gaagaaaatc tattaagctt ccacaaaatt
17041 cggtaatttt tccacatgaa atcaaagatc gtggtgaaga agagagtaag gagaaggaat
17101 ttttccatca aggtaaacaa aatctctaat accttaatta cttccgttta gtaattctcc
17161 ttttacttgt tttttttta atgagagtat gtgacaattt cataaagaaa ttagttgttt
17221 gacatacgag atggttttt gactaattat atttttgtt ttgaaagatt tccaagctaa
17281 ttttaatgag catatttttg attttattga ttgaggaaat tttcagaatt tcgacattta
17341 agtttttttt ttgttttaaa tatacttttg attcgatgat aagagattgg gaaagcagac
17401 taatgatgtt ttgttgtcac gttcattgat tagagatctc ttatattcat atttgtctac
17461 aatatatcat gcatgtgttg atttgtttcg ttaattcaat ttttttttt tcatgttgac
17521 agatggttca caacacactt atcaaaatgg cgagacaaag aattcaaaag agcatgaaag
```

17581 aaagtgtgat gaatcagcac atcttcaagg taaataattt taaattcatt cttaaaaaag
17641 ttagcttatt ggtaagttca ttacaattta tatttaacca tcgtcacttt ttatttaacg
17701 agtttgataa gcattttcaa aacctgtcct tcatctgccg atgcagatgt ggttatgttc
17761 atctttgatt ttattgattg aggatttttt cagaatttcg attcatactt gtctgtaata
17821 tatcatccat gttgttttgt aatcagttaa ttcacttatt ttatttttaa cttttattgt
17881 aacagataat tcacaaacca cccataaaaa aaaggagaag aagaattcaa aagaaaagca
17941 tggaataaag cattctgaat cagaacatct tcaaggtaaa tacttttgaa ttcattcatt
18001 aaaaaaacag tttatttgta agttcattac agtttatata tatttaaatt gtttatgata
18061 atgtattttt gcacaatcga ctaatcatta cccactcatt catttatatt ttatttttatg
18121 gtgaaagatg atatttcgca acgtgttacc ggaaaaggaa ggagaaggaa ttcaaaaggg
18181 acaccaaaaa aactgaggtt taataggcct cggatcttgg aagacggaaa gaaaccaaga
18241 aatcccgcca ccactcgact gagaactata tccaacaaga ggaggaaaaa ggacatagac
18301 agtgaagatg aagttatacc agagcttgca actccaacaa aggaaagctt tccaaagaga
18361 agaaagaacg agaagattaa gagatccgtg gctcggactt taaattttaa gcaagaaatt
18421 gttctgagtt gtcttgagtt cgacaagatt tgtggaccaa tttttccaag agggaaaaag
18481 aggaccacca cacgacgcag atatgatttc ctttgttttt tacttccgat gcctgtttgg
18541 aaaaaacaat caagaaggtc taagcgtagg aaaaatatgg tcagatgggc tagaattgct
18601 tcttcttcaa aactgctaga agaaactttg cctttaatag taagtcatcc gactattaat
18661 ggacaagcag atgcttcttt acacattgat ggtaatcgag ttttttttt gttaatttat
18721 ctgttacatc aaaattgttt atgcttatat ctaaagtatc attgtgtatt atttttgca
18781 gacacactcg tgagacatgt agtctcaaag caaaccaaga aaagtgctaa caatgtcatt
18841 gagcatttaa atcgacaaat aacttatcag aaagatcacg gtctctcatc tctggcagat
18901 gttcctttgc acattgaagg taatctagtc ttatttttgt tcttttttaa tatattgatt
18961 aaaaagattg tgatatattt atttaatata tttttgttat attatatcta tattttattg
19021 tttgtacttt ttttttgtag atacactaat aaaatcggct agttctgtac tttcagaacg
19081 acccatcaag aaaactaagg atattgctaa gttaatcaaa gatatgggaa gattaaagat
19141 caataaaaag gtaacaacga tgatcaaagc tgacaagaaa ctcgttacgg caaaggttaa
19201 tcttgatcca gagaccatta aagagtggga tgtcttaatg gtgaatgatt caccaagccg
19261 atcatatgac gataaggaga cggaggccaa atggaaaaaa gaaagagaga tttttcaaac
19321 ccggatagat cttttcatta accggatgca tcgcttacaa ggtacattat tgttattatc
19381 attattgtta ttatgatcta tttatacttg tattctaaat tagcttacat atatatataa
19441 ggaatccaag tataagtgag tatgctaagt atatgatcat tttttgaaat tatgtttcct
19501 tccatgatgt ttaaatgatt gtcttgcagg caatagaaag tttaaacagt ggaaaggctc
19561 agttgttgac tcagtggttg gagtttttt gacacaaaat actaccgact atctttcaag
19621 gtaaaatctt tgtttaaatt gttaagaaat ttgaaaaact aattcatata atagatgatc
19681 actttgattg tgagtttcta cagcaacgcg tttatgagcg tggctgcaaa atttcctgtt
19741 gatgcaagag aaggtctatc atactatatt gaggaacctc aagatgctaa aagttctgaa
19801 tgtatcattt tatctgatga gtcaatatca aaggtggaag atcatgagaa tactgcaaaa
19861 aggaaaaacg agaaaaccgg tattatagaa gatgagatag ttgactggaa caatcttaga
19921 aggatgtaca cgaaagaagg atctcgtccc gaaatgcata tggactctgt taattggagt
19981 gacgtgagat tatctggcca aaatgttttg gaaaccacca ttaaaaaacg tggacaattc
20041 aggattcttt cagaaagaat attggtaaga aaaacaaaac ttctaatgaa ctttgtgaat
20101 aatttattca aatgatttaa gactaacact ttttttttt tccttgtttt ctcaagaaat
20161 ttcttaacga tgaagttaac caaaatggaa atattgatct ggaatggctt cgaaatgctc
20221 catcacattt agtgaagtat gtttatgttg gtttttatgt tctcatagat ctcattatta
20281 gtaagcgatc ataaactctt tctattattt tatcaggaga tatctgttgg aaatcgaagg
20341 gatagggctg aaaagtgctg agtgcgtacg actgttagga cttaaacatc atgcgtttcc
20401 ggtatgaaaa tattattatg atttttcatt taacatatat tattaatttt tactgataaa
20461 acccatgtgt taatgtgtag gttgacacaa atgttggtcg tatagcagtt cgactaggtc
20521 tggttcctct tgaaccttta ccaaatggag ttcaaatgca tcaactattc gagttatgtt
20581 ttctcattaa tttgattaag aaaatacatt acaagttact aacaactatc tcctatcgat

```
20641 aaacatgaac tcgtttcagg tacccttcaa tggattcgat tcaaaagtac ctttggccac
20701 gattgtgtaa acttccccaa gaaactttgt aagttcaaat gtttttcctc aatttaagaa
20761 gccaactatt tttacgccat ttgaacacat attacctaat tttatttcta aatattttta
20821 cagatatgaa ctacattatc aaatgataac atttggaaag gtgtgcgtta cttttttctt
20881 ttttatatta atgaataaaa taatattgtt ggtttaatca aattttgtca actttaggtt
20941 ttctgcacaa aaactattcc taattgtaat gcatgtccaa tgaagtcaga atgcaaatat
21001 tttgcaagtg catatgtcag gtacaatctt ttttctcttt cctactttga tacttagata
21061 taacttaatt tgttaattcc ataaatatta aagaaaaatc ttagaataat cataaaaaat
21121 aattgctaaa cgtctcagct attttatata ataaattttc taaatattga gagtgaattt
21181 gagttttaat aattacatta tatatataaa tatataatgt tagaattgac aaattgtgtt
21241 tttttttaat agttctaaag ttcttctcga gagtccagaa gaaaagatgc atgagcctaa
21301 tacttttatg aatgcacatt ctcaagacgt tgctgtagat atgacatcaa atataaattt
21361 ggtagaagaa tgtgtttctt ctggatgtag cgatcaagct atatgttata agccactagt
21421 tgagtttcct tcgtccccaa gagcggaaat tcccgagtca acagacattg aagatgttcc
21481 attcatgaat ctttatcagt catatgctag tgttcctaaa attgattttg acttggatgc
21541 attgaagaaa agtgtagaag atgcacttgt aataagtggc aggatgagca gttctgatga
21601 agaaatatca aaagcattag tgattcccac tcctgaaaat gcatgcattc ctatcaaacc
21661 acctcggaaa atgaagtatt ataatcgact aagaactgaa catgtggtgt aagtatcttt
21721 atgtaaatac tgattatacc atataattta tatgcatttt ttgggaatat ataatctaat
21781 acttgttttt tttgcagtta tgtgcttcct gataatcatg agctgctaca cgatgtaagt
21841 atacacatac tttaagctac aaaaaaatgc aactcttttg tataattaat tagaaaatgc
21901 ttttggtttt ttacatatat tatatagttt gagagaagaa aacttgatga tccaagtcct
21961 taccttcttg cgatttggca accaggtata atacaagcat aatttatcat tgttcacata
22021 actataaact aaatttttca ttcgaataat ttttaggtga aacatcatcc tcgttcgttc
22081 caccaaagaa aaagtgtagt tctgatggat caaagctttg caagataaag aattgttcat
22141 attgttggac tatacgagaa caaaactcca acatttttcg cggaacaatt ttggtaaaca
22201 aaatttacaa tttgatattt taacattggt gacttgaaac tcacataaat tcaattgatc
22261 agattccatg tagaacagca atgcgagggg cctttccact taatggaaca tacttccaaa
22321 ccaatgaggc aagcattttt tcttataatt ttttgtctga gtttttactt aatggtttta
22381 aagagaacac aatggtttat ttttccaggt ttttgctgat catgagacaa gcttaaaccc
22441 cattgtcttt cgtagggagt tgtgtaaggg actagaaaaa cgtgcactat attgtggttc
22501 aacagtgaca tctattttta aacttttaga cacaagacgg attgaacttt gcttttggac
22561 aggtaacaaa cataaatata tattaaattt tttgttgaat tatgaagtta aaataactgt
22621 ggaatgttgt gtggtgctgt gcagggtttt tatgtttgag agcatttgat cgaaagcaac
22681 gagatccaaa agagcttgtc cgacgtctac acactccacc tgatgagaga gggccaaagt
22741 ttatgagtga tgatgatata TAGtttcatt ttattctttt tggtctagtt agcaaattat
22801 ttaaacgaac gaatcttttc ttataataac aagcgattca acgattgagt aaatgcacgt
22861 acgtattgtt tcttgattta aatgcatgta cattataatt atttcacaag tggttttcat
22921 atagtagttg tggatgaaaa agaagagagc ccaagagttg gtcttatggt tatgcctggg
22981 ttttggattg gtggcagtgt cattcaaaac cgagtttatg tttctggtgt gaaggtcctt
23041 gagTGAagga tttcaggaac tgtcttaatg cttcttccca ctttgttgtg caacttttat
23101 tttctctttg ttataagcaa gcctatatgt atcaatgata cagtatcatc tattgttcaa
23161 aaaaattgga attaatatct tcttcgtctc aacatctttg ggtcgatcgt tattcgatga
23221 cagtagcaac tagcgagtct cttgtgatat atcctagcca agcgacctca aaactttttt
23281 tacttcgatt gttgtcagta tttctgtttc agacgttttt agcaaaaaag ttctcatggt
23341 gataaaatta ggcttaaaac agtatgactc tgtctttaag actcagtttc agatagtaat
23401 aataaaatta cataaacaaa gagtggtcat agacgtgtat ctgtaagtgt tgtcagagat
```

SEQ ID NO:17

RICE(Oryza sativa) DMT1

>DMTRICE(1DMTRICE);

MQDFGQWLPQSQTTADLYFSSIPIPSQFDTSIETQTRTSAVVSSEKES
ANSFVPHNGTGLVERISNDAGLTEVVGSSAGPTECIDLNKTPARKPKKKKHRPKV
LKDDKPSKTPKSATPIPSTEKVEKPSGKRKYVRKKTSPGQPPAEQAASSHCRSELK
SVKRSLDFGGEVLQESTQSGSQVPVAEICTGPKRQSIPSTIQRDSQSQLACHVVSST
SSIHTSASQMVNAHLFPPDNMPNGVLLDLNNSTSQLQNEHAKFVDSPARLFGSRI
RQTSGKNSLLEIYAGMSDRNVPDLNSSISQTHSMSTDFAQYLLSSSQASVRETQM
ANQMLNGHRMPENPITPSHCIERAALKEHLNHVPHAKAAVMNGQMPHSYRLAQ
NPILPPNHIEGYQVMENLSELVTTNDYLTASPFSQTGAANRQHNIGDSMHIHALD
PRRESNASSGSWISLGVNFNQQNNGWASAGAADAASSHAPYFSEPHKRMRTAYL
NNYPNGVVGHFSTSSTDLSNNENENVASAINSNVFTLADAQRLIAREKSRASQRM
ISFRSSKNDMVNRSEMVHQHGRPAPHGSACRESIEVPDKQFGLMTEELTQLPSMP
NNPQREKYIPQTGSCQLQSLEHDMVKGHNLAGELHKQVTSPQVVIQSNFCVTPP
DVLGRRTSGEHLRTLIAPTHASTCKDTLKALSCQLESSRDIIRPPVNPIGPSSADVP
RTDNHQVKVSEETVTAKLPEKRKVGRPRKELKPGEKPKPRGRPRKGKVVGGELA
SKDSHTNPLQNESTSCSYGPYAGEASVGRAVKANRVGENISGAMVSLLDSLDIVI
QKIKVLDINKSEDPVTAEPHGALVPYNGEFGPIVPFEGKVKRKRSRAKVDLDPVT
ALMWKLLMGPDMSDCAEGMDKDKEKWLNEERKIFQGRVDSFIARMHLVQGDR
RFSPWKGSVVDSVVGVFLTQNVSDHLSSSAFMALAAKFPVKPEASEKPANVMFH
TISENGDCSGLFGNSVKLQGEILVQEASNTAASFITTEDKEGSNSVELLGSSFGDG
VDGAAGVYSNIYENLPARLHATRRPVVQTGNAVEAEDGSLEGVVSSENSTISSQN
SSDYLFHMSDHMFSSMLLNFTAEDIGSRNMPKATRTTYTELLRMQELKNKSNETI
ESSEYHGVPVSCSNNIQVLNGIQNIGSKHQPLHSSISYHQTGQVHLPDIVHASDLE
QSVYTGLNRVLDSNVTQTSYYPSPHPGIACNNETQKADSLSNMLYGIDRSDKTTS
LSEPTPRIDNCFQPLSSEKMSFAREQSSSENYLSRNEAEAAFVKQHGTSNVQGDN
TVRTEQNGGENSQSGYSQQDDNVGFQTATTSNLYSSNLCQNQKANSEVLHGVSS
NLIENSKDDKKTSPKVPVDGSKAKRPRVGAGKKKTYDWDMLRKEVLYSHGNKE
RSQNAKDSIDWETIRQAEVKEISDTIRERGMNNMLAERIKDFLNRLVRDHGSIDLE
WLRYVDSDKAKDYLLSIRGLGLKSVECVRLLTLHHMAFPVDTNVGRICVRLGW
VPLQPLPESLQLHLLEMYPMLENIQKYLWPRLCKLDQRTLYELHYQMITFGKVFC
TKSKPNCNACPMRAECKHFASAFASARLALPGPEEKSLVTSGTPIAAETFHQTYIS
SRPVVSQLEWNSNTCHHGMNNRQPIIEEPASPEPEHETEEMKECAIEDSFVDDPEE

IPTIKLNFEEFTQNLKSYMQANNIEIEDADMSKALVAITPEVASIPTPKLKNVSRLR
TEHQVYELPDSHPLLEGFNQREPDDPCPYLLSIWTPGETAQSTDAPKSVCNSQEN
GELCASNTCFSCNSIREAQAQKVRGTLLIPCRTAMRGSFPLNGTYFQVNEVFADH
DSSRNPIDVPRSWIWNLPRRTVYFGTSIPTIFKGLTTEEIQHCFWRGFVCVRGFDRT
SRAPRPLYARLHFPASKITRNKKSAGSAPGRDDE

SEQ ID NO:18

>DMTRICE novel 723 amino acid NH2 terminus;

MQDFGQWLPQSQTTADLYFSSIPIPSQFDTSIETQTRTSAVVSSEKESANSFVPHNGTGLVERISNDAGLTEVVGSSAGP
TECIDLNKTPARKPKKKKHRPKVLKDDKPSKTPKSATPIPSTEKVEKPSGKRKYVRKKTSPGQPPAEQAASSHCRSELKS
VKRSLDFGGEVLQESTQSGSQVPVAEICTGPKRQSIPSTIQRDSQSQLACHVVSSTSSIHTSASQMVNAHLFPPDNMPNG
VLLDLNNSTSQLQNEHAKFVDSPARLFGSRIRQTSGKNSLLEIYAGMSDRNVPDLNSSISQTHSMSTDFAQYLLSSSQAS
VRETQMANQMLNGHRMPENPITPSHCIERAALKEHLNHVPHAKAAVMNGQMPHSYRLAQNPILPPNHIEGYQVMENLSEL
VTTNDYLTASPFSQTGAANRQHNIGDSMHIHALDPRRESNASSGSWISLGVNFNQQNNGWASAGAADAASSHAPYFSEPH
KRMRTAYLNNYPNGVVGHFSTSSTDLSNNENENVASAINSNVFTLADAQRLIAREKSRASQRMISFRSSKNDMVNRSEMV
HQHGRPAPHGSACRESIEVPDKQFGLMTEELTQLPSMPNNPQREKYIPQTGSCQLQSLEHDMVKGHNLAGELHKQVTSPQ
VVIQSNFCVTPPDVLGRRTSGEHLRTLIAPTHASTCKDTLKALSCQLESSRDIIRPPVNPIGPSSADVPRTDNHQVKVSE
ETV

SEQ ID NO:19

>DMTRICE nucleotide sequence from PAC PO489G09;

```
10261 aaatattgct taaatggata taaagttgaa aaatgtactt gagggaagtt gtaggtgcac
10321 gtggggtccc acaatttttc ttcactagtg caccttttagt tatatatttt ttgcgcaaga
10381 ggacaaaggc gctccgtgta attttgagta agggccggcg ggatatttat ttgtgtaaag
10441 gacctagcca agaaaagcat gatagtgcat atgtatcctt tctttttctt ttcttttgtt
10501 ttcataactg tcttacagaa tttcatgttg gctggtgaca cttgtctcac tcattatttg
10561 gtatattttg actaaatgca acgtgttggt gctcggtagt ttatatttgt ttttacgcat
10621 tcttcattga ctgtatgtat ttgatgttga taccctgggc tgtcttattt tataggtgga
10681 tgctgggagg ccacatagga ggcctgtgtg atccaagtgt gctgctcctg agttgaaatt
10741 gcatagccat atagcaacta ctggtgtaaa cttgagagat gaagtagtga aaggaaatat
10801 gcaggatttt ggacaatggc tgcctcaatc tcagaccact gccgatctat atttctccag
10861 tattccaata ccatcacagt tcgatacttc catagagacg cagactagaa cttctgcagt
10921 tgtatcgtca gagaaagaat ctgctaattc gttcgtccct cataatggta ctgggcttgt
10981 tgaacgcatt agcaatgatg ctgggctaac tgaagtagtt ggaagtagtg ctggaccaac
11041 tgaatgtatt gacttgaaca agacaccagc acggaaaccc aagaagaaaa agcacaggcc
11101 aaaggtgcta aaggacgata aaccatcgaa gacacctaaa tctgctactc caataccttc
11161 aacagaaaag gtagaaaaac catctggaaa gagaaaatat gtccgcaaga agacatctcc
11221 aggccaacct cctgcagaac aggcagctag ctcacactgc agatctgagc tgaagtcagt
11281 taaacgaagt ttggactttg gtggagaagt actgcaagag agtacacaat ctggatctca
11341 agttccggtg gcagaaatat gtactggtcc caagcgtcaa tcaatacctt ctaccatcca
11401 aagagattcg caaagccagt tggcttgcca cgtggtttct agcaccagct caattcacac
11461 ttcagctagt cagatggtta atgcacattt gtttcctcct gataacatgc caaatggagt
11521 attgcttgac ctcaataatt ctactagtca gttacaaaac gaacatgcta aatttgtgga
11581 cagtccggca cgtctttttg gttccagaat aagacagaca tcaggtaaaa attctttgct
11641 agaaatctat gctggcatgt cagatagaaa tgtacctgat ctcaacagtt caatcagtca
11701 gacgcatagc atgtctactg attttgctca atacttgctt tcatcctcac aagcttctgt
11761 aagggaaaca caaatggcca atcagatgct taatggtcat aggatgccag aaaatccaat
```

```
11821 tacacctagt cattgtattg aaagggctgc attgaaggaa catttgaatc atgttcctca
11881 cgcaaaagcc gcagtgatga atggccaaat gccccatagt tacaggttgg cgcaaaatcc
11941 catcctacct ccaaatcata ttgaagggta tcaagtgatg gaaaatttga gtgaacttgt
12001 cacgacaaat gactatctaa ctgctagtcc tttcagtcaa actggagctg caaataggca
12061 gcataatatt ggtgactcca tgcatataca tgcattggat cctagaagag agagtaatgc
12121 ttcaagtggt tcttggatat cattaggtgt gaactttaac caacaaaata atggatgggc
12181 atctgcaggt gctgccgatg ctgcgagctc acatgcccca tatttttcag aacctcacaa
12241 aagaATGagg acagcttatc ttaacaatta tccaaatgga gtcgtgggac atttttctac
12301 ctcatctacg gatttgtcaa ataatgagaa tgaaaatgtg gcctcagcaa tcaactcaaa
12361 cgtttttacc cttgctgatg cacaaagatt gatagcccgt gagaaatcac gagcttccca
12421 aagaatgatc agttttagat catctaaaaa tgatatggtt aacagatcag aaatggtcca
12481 tcaacatggc agacctgctc cgcatggctc tgcatgcagg gagtctattg aagtacctga
12541 caaacagttc gggctcatga cagaagaact cacacaatta cctagtatgc caaataaccc
12601 acaaagggaa aaatatattc cgcaaactgg aagttgccaa cttcagtctt tggaacatga
12661 catggttaaa gggcataact tggcaggtga attgcataag caagtaactt cacctcaagt
12721 tgttattcag agcaatttct gtgttacccc tcctgatgtg ctcggcagaa gaaccagtgg
12781 ggagcattta agaaccctta tagctccaac acatgcatcg acatgtaagg acactctgaa
12841 agctttaagt tgtcaactgg agagttctag agacattatt aggcctcctg tcaatcctat
12901 agggccatcc tctgccgatg ttccaagaac tgataaccat caagtcaagg tttctgaaga
12961 aaccgttaca gccaaactcc ctgagaagcg aaaagtagga cgtcccagaa aagagttaaa
13021 acctggtgag aaaccaaaac ctagaggccg tccaaggaag ggaaaagttg ttggtggaga
13081 acttgcatca aaggatagtc acactaatcc attgcaaaat gagagtactt catgttctta
13141 tggtccttat gcaggggagg cttctgttgg aagagcagtt aaagcaaata gagttggaga
13201 aaacatttct ggagctatgg tatccctact ggattcttta gatattgtta ttcaaaagat
13261 aaaggtcttg gacataaaca aatcagaaga ccctgtgaca gctgaacctc atggtgctct
13321 tgtcccttac aatggagaat ttggtcctat tgttcctttt gaggggaaag tgaaaagaaa
13381 acgctctcga gccaaagtgg atcttgaccc tgtaactgct ttaatgtgga agttactaat
13441 gggaccagat atgagtgatt gtgctgaagg tatggataag gataaagaga aatggctaaa
13501 tgaagaaaga aaaatattcc aagggcgtgt tgattcattt attgctcgaa tgcatctagt
13561 tcaaggtatt tctatcattt taaaattgtt ttcctaacat gaacatgatg gcttccatct
13621 tgtgattgct gccctcacat tagtgaatgg tctcaaatct tcaatattta ctgtgtaccc
13681 aaatcctatt tcttcatccc aatatattca tgtttgtact cgtactgtcc cattagactt
13741 gcattgtgct gtgaagatca acacctttac ttttaggatt acctctatgt ttgcaggaga
13801 tcggcgtttt tctccttgga aaggatcagt tgtagattct gtagtgggag tatttcttac
13861 acagaatgtt tcggaccatc tttccaggtg aataatgcct agagcctatt tgaaaactgt
13921 gacttgactt gcattgtgag gttatgttgt ttttctgtct gactatttcc tttttttca
13981 gctctgcatt tatggctctt gctgcaaaat ttcctgtaaa gccagaagcc tctgaaaaac
14041 ccgcaaatgt gatgtttcat acaatttcag aaaatggtga ttgttctggg ttgtttggta
14101 attctgtcaa gctacagggt gagatccttg ttcaggaggc cagcaacaca gcagcctctt
14161 ttatcacaac cgaggataag gaaggaagta acagtgtgga attgcttgga agttcttttg
14221 gggatggagt ggatggtgca gcaggagttt attctaatat ttatgagaat ctgccagcta
14281 gactgcatgc tactaggcgt ccagtcgttc aaactggaaa cgctgtcgaa gcggaagatg
14341 ggtcactgga gggtgttgtt tcatcagaaa actccactat ttcatctcaa aattcatcag
14401 attatctatt tcacatgtct gatcatatgt tttcgagcat gttactaaat ttcactgccg
14461 aagacattgg cagcagaaat atgcccaaag caacaagaac cacatataca gaacttctac
14521 gaatgcagga gctgaagaac aagtctaatg aaaccattga atcatcagag tatcatgggg
14581 ttccagtctc atgtagtaac aacattcaag tgctcaatgg aatacaaaat atcggcagta
14641 aacatcagcc tttacattcc tctatttcat atcaccagac tggccaagtt cacctcccag
14701 acatagtaca tgcgagtgat ttggagcaat cagtatacac tggccttaat agagtgcttg
14761 attctaatgt tacacaaacc agttattatc cttcacctca tcctggaatt gcctgtaaca
14821 atgaaacaca aaaggctgac tctttaagca acatgttata tggtatagat agatcagata
```

14881 agactacttc cctgtctgag cctacaccaa gaatcgataa ctgttttcaa ccattaagtt
14941 cagagaaaat gtcatttgct agggaacagt cctcttctga aaattatctt tcaaggaatg
15001 aagctgaagc tgcatttgtt aaacagcatg gaacatcaaa tgtgcaaggt gataatactg
15061 tcaggacaga gcaaaatgga ggtgaaaatt ctcaatcagg atacagccaa caggatgata
15121 atgttggatt tcaaacagcg acaaccagta atctttattc ttcaaactta tgccaaaacc
15181 agaaagcaaa ttctgaagta ctacacggag tttcttccaa cttgatagag aattctaaag
15241 atgacaaaaa gacttccccc aaagttccag tcgatggatc aaaagcaaag aggccaagag
15301 ttggggctgg taaaaagaaa acatatgatt gggatatgtt gagaaaagaa gttctttaca
15361 gtcatggtaa taaagaaaga tcccagaatg ctaaggactc aattgattgg gaaacaataa
15421 gacaagcaga ggtgaaggaa atatctgaca caattagaga gcgaggaatg aataacatgc
15481 tggcagaacg gataaaagta agtatggcat aaaacagttt acattgaaag ttgacataac
15541 tctagtcata tgtgcatgca tgctattcca tatagatttg cttatttgtt ggaattccaa
15601 gttttggatc aaccatactc atctttagca attcatgttg caggacttcc taaaccgatt
15661 ggtgagagac catgggagca tcgatcttga gtggttgcgc tatgtcgatt cagataaagc
15721 gaagtaagct aactaaattt attttgagca aacattcata atgcaattgg cccttgggca
15781 ttctataatt tgtcattttg acctctgcat tgcttagcaa tgacaattgg atgtagtgag
15841 catgggtaat aatgtaagca atgacaattg gatgtagtgg gcatggttaa taattgaaca
15901 tgtctgtgtt tgcgggataa taatgcctat cacctgtgag cctgtgacat gcaaaccttg
15961 aacgttgaac cttgaacccc ctacctcgca ctgtgtgctc tcaaccaact gagcaagtga
16021 gggaccttgt tgtatggaaa aaataatttt aaataaccct tgattcaacc aaagcttcat
16081 aaaagaatat attttctatt attcatttga accagcggtt gaaccagtga accgatggtc
16141 ttgctggtcc ggatttaata ataactatgg ctagaacaga ttagagcacc gaatacttgc
16201 gcgatgctaa atatttcaat ggggacacac ctgctcgtgt gttgcatcaa ctacctaagc
16261 cacacaggca tggcaatcaa atcagcttgc ccatgtaaca tcaactatct gatcgcgaga
16321 aggccggagc tctcacttga tgtttgtcat tcaaaaaata gttattcacc aatgcaatgt
16381 caagctcccg taaagaccat gaatgtagtt tatccttctt tgatcaagtt tttatttata
16441 ttaaagtgtt taccaatgta atcctacatt atttgtacct ggtttttaca tataaataca
16501 ttgtaccttt tgtgtttctt ccagggacta tctcttaagc attagaggac ttggacttaa
16561 aagtgttgag tgtgtgcgtc ttttgacact ccatcacatg gcttttcctg tatgtttcct
16621 ttcacaaata attttcaaga atcttcgttt ctttatttct ggagaagtgg agattttatc
16681 tgtatctgtt gatgatgtag gtggatacaa atgttggtag aatatgtgtg aggcttggat
16741 gggtgccact tcagccccta cccgagtctc ttcagttgca cctgttggag atgtaagtat
16801 cttaaatcca ctggttggct tcactaatgc tggagagtga taggagtttg atcatctgct
16861 attgaaggta tccaatgctg gagaacatac agaaatacct ctggccgagg ttatgcaagc
16921 ttgatcaacg gacattgtga gttttagaaa tgcagttaaa aactatatat ataagagcat
16981 gtcattatct gagagtgtaT AAcaggttct tgatgatatg taggtatgag cttcactatc
17041 aaatgataac ttttggaaag gtatgagaca acaactttga taaagtgaat tcaacccaat
17101 tactgtgttt tgatggacca tctgtgttac tttccttcta ggtattttgt acaaaaagta
17161 agcccaattg caacgcatgc ccaatgagag ctgagtgcaa gcactttgca agtgcatttg
17221 ccaggtaatt ctcaagatgt acatatttta tatacattct gtgaaatcac ggtgatgatt
17281 gttaggtatg aacaattggc tgagatcccc ccctccccc ctcccatcct tttcctggtc
17341 ctacaagttc tcctaggcta atttaactgg tgcataccac atttatgtta ttttgataca
17401 tcaaagatta tgtttgtggt tgtgaggcta tattagtgtg ttgtatgtaa ctcagttttg
17461 caattgtagt tttagttaga acacgttgtt ctctacattt taataaatac tttttgactg
17521 gacatcaatg actggtgtat ttccgatata aaaaggttga ttgttgccga gggatttcaa
17581 ttcggtccga ataggttcga caaatgcagt gggcctatta gtttaagagt gaaagttcta
17641 tcagctgttt gactccactg tgacctttac actttgtact tttgaagaaa cagactaacc
17701 tgctcatatt aaagtcttgg aatgactcca ttgcgacctt tacgctttgt attttagaag
17761 aaacagacta acctgttcat attagagtct tggaactgtg tgtgtgtgtg ttttttttt
17821 ttttgggggg gggggggcat ggagatttaa tccaacattc ctggatgacc ttatattggt
17881 aatgatatgg ttttttttatg atatagtgca aggctcgctc ttcctggacc tgaagagaag

17941 agtttagtta catctggaac cccaatagct gcagaaacct tccaccagac atatataagt
18001 tctaggcctg tagtaagtca gcttgagtgg aattcaaaca cctgtcacca tggtatgaac
18061 aatcgccagc caatcattga ggagccagca agcccagaac ctgaacatga gacagaagag
18121 atgaaagagt gtgcaataga ggatagtttt gtcgatgatc cagaagaaat ccctactatc
18181 aagcttaatt ttgaggagtt tacacagaac ctgaagagtt atatgcaagc aaataacatt
18241 gagattgaag atgctgatat gtcaaaggct ttggtcgcta taactcctga agttgcttct
18301 atcccaactc ctaagctcaa gaatgtcagt cgcctaagga cagagcacca agtgtatgat
18361 cttgtccctc ttgcaaaacc aatctcatga atatttacta ttgactatca tgtgttttgc
18421 tgcattgctt acttctctgt tttcaacata tatgtagcta tgaactgcca gattcacatc
18481 cacttcttga aggagtaagt tcataaaaca ttatagaatt ctgtactttc cttatcacca
18541 actgagaata tattgatgct tattttctta caatacacag ttcaaccaaa gagaaccaga
18601 tgatccttgc ccatacctac tctctatatg gaccccaggt aagaagtgca taaacagaac
18661 acaatatcat gggaaccaaa ctttttttcaa tggttactta taattgttga aatatgcaac
18721 aggtgaaaca gctcaatcaa ctgatgcacc taagtcggtc tgcaattcac aagagaatgg
18781 tgaactatgt gcaagcaata catgctttag ttgcaacagt ataagagaag cgcaggccca
18841 aaaagttcga gggacactgc tggtaagtag ttgtttctgt aacatatgct cagttgccct
18901 tggttcaaga tgtgctattc aagtttatca tgttcacgaa tagtgataaa gctgctatct
18961 gtcctagcta ttgtccaagc tataacagtt ctgattcact ggttgggcac cagctaggga
19021 ataggatgta aaaaacttat cccgcagttt gttgacaatc tgtttttctt tgttgaaaat
19081 taaaaatага taccatgccg aacagcaatg agaggaagct ttccacttaa tgggacatat
19141 tttcaagtca atgaggtgaa aacagaaagt tcttaaagtt gatcttagtt taattattat
19201 aataccatta aaatatatgc aagtttctac tttctagtat ctcttttatt agtgttcaaa
19261 tgttatgcgg caggtatttg ctgatcatga ctcaagccgg aacccgattg atgttccaag
19321 gagttggata tggaatctcc ctaggagaac tgtttacttt ggaacttcaa ttccgacaat
19381 atttaaaggt atttcactaa taaattttga ccaagaatag gatttttggc agcgccaaat
19441 gtgccactat ctttattgtg tgaagtccat tatgtgattg taataatttg aatcaccaag
19501 aggactaagg cctgctttgg gacatattac gagcagcttt tgcttgcaaa gaaaccagat
19561 tctggtgccg caccttctcc gctcttctgc cacccaagtc cgtccaatac ccctcattga
19621 gcgcttggat cctaacccca tctgccatca tgcatcatcc tgctaacaac tgcttccacc
19681 attgcctgtt tctgttgttg ggaggcactc acgctgcttg ctatagttta ggttttcttt
19741 gtgtcctgat ttagatggaa tttccagctg ctgtctttta cataactagc taaatgtccg
19801 cgctttgcta tggataatag aaaatatatt ataatattgt caaataaatt aaatatgttt
19861 tatacgaaat gtgttaacaa tccttttgct atagggaata ttgaccttaa tttgatttta
19921 tatgtggcta tccatttaga tttgtttgtt tttctaataa taataagttc aagggctaat
19981 gtacaaaatt gacaatggga gtaggtgggg tggcagattc actgccacca ccactacctt
20041 cttttaaagg ggtatataga tttgcagcag tggttgcttg atctgtgatt tgaaatgtca
20101 agtacacgct catgcatcag caccatatgt ctacgctcct gacccaacat gcaaccaatg
20161 caattgaggg ttggctctga tacaattact aatgtcctat atccaaaaca actataggcc
20221 tatgaccaaa cataattaat aacctcgctt gcgcttttgt cctcacttgc tccatgtaaa
20281 agggttaacc cgaggttact atgttaggaa tagctgggtt tatgaaacgg ttcaactctc
20341 aactcctcat atagcactaa ttcatgtatt gctgtcagca gtgatttgag ttccagatca
20401 tgctcataag ataggaccaa attgtcctta ctatctactc cctccgtccc aaaatataag
20461 gtatttccgg tcaaaatatc ttatattttg ggatggaggg agtactatac tacggaccca
20521 ccaccaaata gtgccgcaga agagagagag agagagagaa gaggggtgg gggtgggggt
20581 gtatgggtga aataagaata gtgccaagta tttgccaaca aatgaggcgg tcaaatgtgt
20641 cacatcaatt gggaagtatg tcagatcaac tgaaaatttg attgggaaat tattattcat
20701 gcaacaaagc tgtacaactg atcccatgtt tctatcgcag gtttgacaac tgaagaaata
20761 caacattgct tttggagagg taatcatttt tttttgtatg tacgttttgg tttccataac
20821 aaagagagat gaagtgtata ggtactatgt ttactgacaa ggataataat agtagcaagt
20881 atataggcag aggagcatgt ctctattcta ccagtattat tactcataat aactagtata
20941 tccttttttt tgccatttca gctgatagct actctccagt caaaatattt gccatctcta

```
21001 ttgaactttt cattgtcttc tgaatgtatc ttactcttgg atcattaata tttcattttg
21061 tcacgatata gtggtatagg acaataaaat catgggaagt atttattttc atcaccaatc
21121 tactcatata attttcaaat gacaattata aatatcttaa aaatatattg ttagttgtcc
21181 tgtataaaat aattgtcaca ccctagtcca cagcgacaag aatttgtgtc tacaggctag
21241 agtgagtact ctagaagtat cttcatagga atcggaataa aatgccaatg tgaatgaaca
21301 aggatatcaa gtataccctc aaaatctcta gagaggattg cgtaaatatg taggtgtaat
21361 taaacaattg tttcatatgg agggttttct taaaggaggt acaagactta tcaatatggg
21421 taaagtagtt tttatccata ggcattgttg gcagaaagct gcttagggta gaatgctact
21481 ccctccgtcc cacaatataa gagattttga gtttttgctt gcaacgtttg accactcggc
21541 ttattcaaaa atttttgaaa ttattattta ttttatttgt gacttacttt attattcaca
21601 gtactttaag tacaactttt cgttttttat atttgcaaaa aaaattgtat aagacgagtg
21661 gtcaaacgtt gtacgcaaaa actcaaaatc ccttatattg tgggacggag ggagtactta
21721 tggatgcctt ttttgtccaa gatgtcagta acattttctt tcagggatgt ggatttttac
21781 ttctttttc cctaactttt tcaggatttg tgtgcgtgag aggctttgat aggacatcaa
21841 gagcacccag accactgtat gcaagactcc actttccagc aagcaaaatt accaggaata
21901 aaaaatctgc aggttctgct ccaggaagag atgatgaata ggccatctgg aaaaccagaa
21961 aggaaataaa gaggaggtac atatgatctg ccagaagatc actgacctga aatggatcgc
22021 tgaccaataa gttgccgtag gcaattcaat tatttctggc catatacatc tgctgaaagt
22081 tatgaactcc agccactgac gaattcgtgg tgctggtatt cttcggcaac atgatccatc
22141 atacagattc tatgcttggt tgttgcaagc aattcttatg cggtgacagt tgctgctgat
22201 agggagaaaa ggcatgtccg gcggctcagc ggctctaact gtactttcat atgagtggaa
22261 ccgattgttg tacatgtgaa aagtttgcca ttcaaaatgg tcattcatgt tgttaggtca
22321 ttcatgtagt cgatgtcaaa ttaatcatca attatttgat ttgattcatt cacaagttta
```

SEQ ID NO:20

CORN(*ZEA MAYS*)DMT.1

>Corn DMT.1 660990 (688512 selclone ID);

EPDDPCPYLLSIWTPGETAQSIDAPKTFCDSGETGRLCGSSTCFSCNNIREMQAQKVRGT
LLIPCRTAMRGSFPLNGTYFQVNEVFADHCSSQNPIDVPRSWIWDLPRRTVYFGTSVPTI
FRGLTTEEIQRCFWRGFVCVRGFDRTVRAPRPLYARLHFPVSKVVRGKKPGAARAEE

SEQ ID NO:21

>Corn DMT.1 cDNA 660990 (668512 selcone ID);

gaaccagatgatccttgtccatatcttctttccatatggaccccaggtgaaactgcacaa
tcgatcgatgcccccaagac
attctgtgattcaggggagacgggtagactatgtggaagttcaacatgctttagttgcaa
caatatacgagaaatgcagg
ctcagaaagtcagaggaacacttttgataccatgccgaacagcaatgagaggaagcttcc
cacttaatgggacgtatttt
caagttaatgaggtatttgctgaccattgctcaagtcaaaatccaattgatgtcccacga
agttggatttgggacctccc

aagacgaactgtttactttggaacctcagttcctacaatattcagaggtttaacgactga
agagatacaacgatgctttt
ggagaggatttgtttgcgtgaggggctttgataggacagtgcgggcaccaaggccccttt
atgcaaggttgcattttcct
gtcagcaaggttgttagaggcaaaaagcctggagcagcaagagcagaagaataatagaac
attgaagaaatataggggtg
ctaaccagatgaggatggatagcccgaaatgagatgctgacccaataggtcgccaaatca
cctccaaattctaacccaat
gacttccatctgtaatgaatggcaataccttgaaaacctgtgatggagatgttttgtggc
gacatgatctcttaaattag
attccgtctttggtaacagcctagctgttcttgttgagtcgcatattctttattctgaag
atcaatatagcaaatggg

SEQ ID NO:22

CORN(*ZEA MAYS*)DMT.2

>Corn DMT.2 371537 (441428 selclone ID);

MITFGKVFCTKRQPNCNACPMRSECKHFASAFASARLALPAPQEESLVKLSNPFAFQNSS
MHAMNSTHLPRLEGSIHSREFLPKNSEPIIEEPASPREERPPXTMENDIEDFYEDGEIPT
IKLNMEAFAQNLENCIKESNNELQSDDIAKALVAIXTEXASIPXPK

SEQ ID NO:23

>Corn DMT.2 cDNA 371537 (441428 selclone ID)

tatcagatgattacatttggaaaggtcttttgtaccaaaagacagccaaattgcaatgca
tgcccaatgaggagtgagtg
caagcattttgcaagtgcatttgcaagtgcaaggcttgcacttcctgctccccaggagga
aagcttagtgaagttgagca
atccatttgctttccagaatagcagcatgcatgctatgaattcgactcacctacctcgcc
ttgaggggagtatccattca
agggagtttcttcctaagaactcagagccaataatcgaggagcctgcaagtccaagagag
gaaagacctccakaaaccat
ggaaaatgatattgaagattttatgaagatggtgaaatcccaacaataaagcttaacat
ggaagctttgcacaaaact

tggagaattgcattaaagaaagcaataacgaactccagtctgatgatattgcaaaagcat
tggttgctattarcactgaa
rcagcttcsattcctgkaccgaaact

SEQ ID NO:24

Corn(*Zea mays*)DMT.3

>Corn DMT.3 218853;

MPRKPKRKAPASPARHDPSPEPYPSHASPCSAQCLVVRDALLAFHGFPEEFAAFRVLRLG
GLSPNRDPRPSSPTVLDGLVTTLLSQNTTDAISRRAFASLKAAFPSWDQVVDEEGKRLED
AIRCGGLAATKAARIRSMLRDVRERRGKICLEYLRELSVDEVKKELSRFKGIGPKTVACV
LMFYLQKDDFPVDTHVLRITKAMGWVPATASREKAYIHLNNKIPDDLKFDLNCLFVTHGK
LCQSCTKKVGSDKRKSSNSACPLAGYCCIGEKLQQL

SEQ ID NO:25

WHEAT DMT.1

>Wheat DMT.1 614028 (887053 selclone ID);

MRAECKHFASAFASARLALPGPEEKSLVTSGNPIASGSCQQPYISSMRLNQLDWNANAHD
HILDNRQPIIEEPASPEPEPETAEMRESAIEDIFLDDPEEIPTIKLNFEEFAQNLKNYMQ
VNNIEMEDADMSSALVAITPEAASIPTPRLKNVSRLRTEHQVYELPDSHPLLEGYDQREP
DDP

SEQ ID NO:26

>Wheat DMT.1 614028 (887053 selclone ID);

tgcccaatgagagctgaatgcaagcactttgcaagtgcatttgcaagtgctagacttgctcttcctggacctg
aagagaagagtttggttacgtcaggaaacccaattgcttcagggagctgccagcagccatacataagttctatgcgtttaaatcaa
cttgactggaatgcaaatgcccatgaccatattctggacaatcgccagccaatcattgaggagccagcaagtccggaaccagaa
ccagagactgcagagatgagagagagtgccatagaggatatttttcttgatgatcctgaagaaattcctacaatcaagcttaatttc
gaggagtttgcacagaatctcaagaattatatgcaagtcaataacattgaaatggaagatgctgatatgtcaagtgccttggttgcc
ataactccggaagctgcatctatcccgactcctaggctcaagaatgttagtcgcctaagaacagagcatcaagtctatgaactgcc
ggactcacatccacttctggaaggatacgaccaaagagagcctgatgatccttg

SEQ ID NO:27

Wheat DMT.2

>Wheat DMT.2 568842 (908118 selclone ID);

NRVDESTVGGADKAASPKKTRTTRKKNTENFDWDKFRRQACADGHMKERKSERRDSVDWE
AVRCADVQRISQAIRERGMNNVLSERIQEFLNRLVRDHGSIDLEWLRDIPPDSAKDYLLS
IRGLGLKSVECVRLLTLHHLAFPVD

SEQ ID NO:28

>Wheat DMT.2 568842 (selclone ID 908118);

caaacagggtggatgaatctactgtcggaggagcagataaagcagcaagtccaaagaaaacaagaaccacaagaaaaaaa
aatactgaaaacttcgactgggacaaatttcgaagacaggcctgtgctgatggccacatgaaagaaaggaagtctgaaag
aagagactctgttgattgggaagcagtacgatgtgcagatgtacaaagaatttctcaggccatccgggaacgaggaatga
ataatgttttatcagaacgaatccaggaattcctgaatcgcttggttagagatcatggaagcattgatcttgaatggtta
agagatatccccctgactcagcaaaggactacttgcttagcatacgtggactggggctcaaaagtgttgaatgtgttcg
tctactgacattacatcatctcgctttccctgtwgacac

SEQ ID NO:29

WHEAT DMT.3

>Wheat DMT.3 611792 (838515 selclone ID);

NRKQVNEVFADHKSSYDPIYVAREQLWKLERRMVYFGTSVPSIFKGLTTEEIQQCFWKGF
VCVRGFERETGAPRPLCQHLHVAASKVPRSRNAAAAGLNSDSAKASAP

SEQ ID NO:30

>Wheat DMT.3 611792(838515 selclone ID);

aatcgaaaacaagttaatgaggtatttgcagaccacaaatctagctacgatcccatatacgttgcaaggga
gcagttatggaagttggaaagacgaatggtctactttggaacttcagtgccctccatattcaaaggtctaacaactgaagaaataca
gcagtgcttctggaaaggatttgtctgtgtgcggggattcgagagggaaaccggggcaccaaggcctctatgccaacatctgca
cgtcgcggctagcaaagtgccgagatcacgcaacgcggcagcagctgggctgaactcggattcagcaaaggcatcggctcca
tgagtatcatcacaccggctatcgacctgtgcatgggtacgctagtgttggttcctgccgggcwacagccgttyttgtaggaaata
aaccsctgcgcaaragaattatcatccagttggtytgagtgtatacttytgctgtagkacctttttTaaaatccctgtgagctytattg
taccttgaatttactttccgaccagtttatccgcttgcaaaraggcctttgttatgkaccggcatcttgttgtatatacatcatggttcctc
traaaaacttgtcttgccakacgaccttacgt

SEQ ID NO:31

Wheat DMT.4

>Wheat DMT.4 615131 (861906 selclone ID);

MRSECRHFASAFASARLALPAPQEKSLVMSSNQFSFQSGGMPTPYSTVLPQLEGSAQGQD
FCTNNSEPIIEEPASPAREECPETLENDIEDYDPDTGEIPLIKLNLQAFAQNLENCIKES
NMDLGSDDIAKALVAVSTGSASIPV

SEQ ID NO:32

>Wheat DMT.4 615131 (861906 selclone ID);

tacttttggaaaggtgttctgtacaaaaaacaagccaaattgcaatgcttgtccaatgag
aagcgaatgcaggcatttcgcaagtgccttcgcaagtgcacggcttgcacttcctgcacc
tcaggagaaaagtttggtgatgtcgagcaatcaattcagtttccagagtggtggcatgcc
aactccatactcaactgtgcttcctcagcttgagggaagtgcccagggacaggatttttg
cactaacaattcagagccaattattgaggagccagcaagtccagcacgggaagaatgtcc
agaaactcttgaaaatgata
ttgaagattacgatccagatactggtgaaatcccactaattaagcttaacttgcaagctt
ttgctcagaacttggaaaactgcattaaagaaagcaatatggatcttgggtctgatgata
tcgcgaaagcacttgttgctgttagcactggatcagcttcaattcctgtccc

SEQ ID NO:33

Soybean(*Glycine max*)DMT.1

>Soy DMT.1 449122 (557119 selclone ID);

MDSLDWDAVRCADVSEIAETIKERGMNNRLADRIKNFLNRLVEEHGSIDLEWLRDVPPDK
AKEYLLSIRGLGLKSVECVRLLTLHHLAFPVDTNVGRIAVRLGWVPLQPLPESLQLHLLE
LYPVLESIQKYLWPRLCKLDQETLYELHYQMITFGKXFCTKSKPNCNACPMRXECRHFAS
AFASARFALPGPEQKSIVSTTGNSVINQNPSEIISQLHLPPPENTAQEDEIQLTEVSRQL
ESKFEINICQPIIEEPRTPEPECLQESQTDIEDAFYEDSSEIPTINLNIEEFTLNLQN

SEQ ID NO:34

>Soy DMT.1 449122 (557119 selclone ID);

aataaaatttaakagcaaggaacaagaaaaagagaaaaaggatgaytttgactgggatagtttaagaattg
aagcacaggctaaggctgggaaaagagaaaagacagataacaccatggattctttggactgggatgctgtgagatgtgcagat
gtcagtgaaatcgctgagaccatcaaagaaaggggcatgaacaacaggcttgcagatcgtattaagaatttcttaaatcgattggt
tgaagaacatggaagcattgaccttgaatggcttagagacgttccacctgacaaagcaaaagaatacttgctcagcataagagga
ttgggactaaaaagtgtggaatgtgtgcggcttttaacactgcaccatcttgccttcccggtagacacaaatgtcggacgtatagca
gtacgactgggatgggtccctctacagccactgcctgagtcactgcagttgcatctcctagaattgtacccagtgttggagtcaata

caaaaatatctctggcctcgactatgcaagctagatcaggaaacactatatgagctacattaccagatgattacatttggaaaggkc
ttctgtacaaaaagcaaaccaaattgtaatgcatgcccaatgagaggagaat

SEQ ID NO:35

SOYBEAN(*GLYCINE MAX*)DMT.2

>Soy DMT.2 387990 (473695 selclone ID);

MRMTIDLVSQQSLTARLQLSILKDKLKIQCRKARGLDFGRNESSKIDSSPVKLRSREHGK
EKKNNFDWDSLRIQAEAKAGKREKTENTMDSLDWDAVRRADVSEIANAIKERGMNNMLAE
RIQSFLNLLVDKHGGIDLEWLRDVPPDQAKEFLLSIRGLGLKSVECVRLLTLHHLAFPVD
TNVGRIAVRLGWVPLQPLPESLQLHLLELYPVLESIQKYLWPRLCKLDQRTLYELHYQLI
TFGKVFCTKSK

SEQ ID NO:36

> Soy DMT.2 387990 (473695 selclone ID);

gaaaagataggatcattctcagatagcaactcagaaatagaagacctgtctagcgctgcc
aagtacaatagttattataa
tagaatttctttcagtgagcttttagaaatggcaagttcaaccatgttgcatgaagttaa
cagtcaaagaagcaaatcaa
ctgagaacttaggagatacatgtgatcagtctatagacatgaagcatgacaacctggcag
aaaacttggaaaaatcggat

gttactcaaggctccgcagaagcacccatcaccaatggatatacttttaaaataacccca
aactcaggagtacttgaggt

taactgttatgatcctctcaaaatagaagtcccatcaagtggctcctcaaagggtaaaga
tgagaatgacaatagatcta
gtttcccaacagagtctgactgccaggctgcaattgtccattctcaaggacaaactgaag
atccaatgcagGaaagcaag
gggactagattttggtaggaatgaaagcagtaagatagattcttcccctgtaaaattaag
gagcagggagcatggaaaag

agaaaaagaataactttgattgggatagtttaagaatacaagcagaagctaaggcaggga
aaagagaaaagacagagaac
accatggactccttggactgggatgctgttagacgcgcagatgtcagtgaaattgccaat
gcaatcaaagaaaggggcat

gaacaacatgcttgctgaacgtattcagagtttcctgaatctattggttgacaagcatgg
gggcatcgatcttgagtggc
tgagagatgttccacctgatcaagcaaaagaattcttgctcagcataaggggattgggat
tgaaaagtgtggagtgtgta
cgactcttaacactacaccatcttgcctttccggtggacacaaatgttggacgtatagca
gtaagattgggatgggtgcc
tctccagccactgccagagtcactacagttgcatcttctagaattgtacccagtgttgga
gtccatacaaaaatatctct
ggccccggctctgcaagctagaccaaagaacattgtatgagctgcattaccagctgatta
catttggaaaggtcttctgt
actaaaagcaagcc

SEQ ID NO:37

SOYBEAN(*GLYCINE MAX*)DMT.3

>Soy DMT.3 657152 (546665 selclone ID);

INQAELQQTEVIRQLEAKSEINISQPIIEEPATPEPECSQVSENDIEDTFNEESCEIPTI
KLDIEEFTLNLQNYMQENMELQEGEMSKALVALHPGAACIPTPKLKNVSRLRTEHYVYEL
PDSHPLLNGWNKREPDDPGKYLLAIWTPGETABSIQPPESKCSSQEECGXLCNENECFSC
NSFREAXFXDSXRDTPDTMSNSXXXGAFH

SEQ ID NO:38

>Soy DMT.3 657152 (546665 selclone ID);

tataaaccaagcagaacttcaacaaacagaagtgatcaggcaactagaagcaaaatctga
aatcaacatcagccaaccta
ttattgaagagccagcaactccagagccagaatgctcccaagtatccgaaaatgatatag
aggataccttcaatgaggaa
tcatgtgaaattcccaccatcaaactagacatagaagagttcactttgaacttacaaaac
tatatgcaagaaaacatgga
acttcaagaaggtgaaatgtcaaaggccttggttgctctacatccaggtgctgcatgcat
tcctacacccaagctgaaga
atgtgagccggttgcgaacagagcattatgtttatgaactccctgattcacatcccttc
tgaatgggtggaacaagcga
gaacctgatgatccaggcaaataccttctagctatatggactccaggggagacagcagat
tctatacagccaccagaaag

caaatgcagctctcaggaatgtggccggctctgtaatgagaatgaatgttttcatgcaa
cagtttccgtgaagcaaggt
tcacagatagttcgagggacactcctgataccatgtcgaacagctwtgaragggag

SEQ ID NO:39

SOYBEAN(*GLYCINE MAX*)DMT.4

>Soy DMT.4 432980 (663678 selclone ID);

EAASIPMPKLKNVSRLRTEHCVYELPDTHPLLQGWDTREPDDPGKYLLAIWTPGETANSI
QPPESKCSSQEECGQLCNENECFSCNSFREANSQIVRGTLLV

SEQ ID NO:40

>Soybean DMT.4 432980 (663678 selclone ID);

agaagctgcttccattcctatgcccaagctaaagaatgtgagccgattacgaacagagca
ttgtgtttatgaactcccag
atacgcatcctcttctccaagggtgggacacacgagagcctgatgatccaggcaaatatc
ttcttgctatatggactcca
ggtgagacagcaaattctatacagccaccagaaagcaaatgcagctctcaagaagaatgt
ggccaactctgtaatgagaa
tgaatgtttctcgtgcaacagtttccgtgaagcaaattctcagatagttagagggacact
cctggtctgaatgcttatca
aaatcattgttttaaccatatgtagcttactaattcttatacattatgggaacaggggag
ggaatacatctccatagaaa
ttcaagcattataatagactgacttgaatttatgataaatatgagcagataccatgt

SEQ ID NO:41

>Medicago 6654943;

MELQEGEMSKALVALNQEASYIPTTKLKNVSRLRTEHSVYELPDSHPLLEGWEKREPDDP
GKYLLAIWTPGETANSIQPPDRRCSAQDCGQLCNEEECFSCNSFREANSQIVRGTILIPC
RTAMRGSFPLNGTYFQVNEVFADHESSLNPISVPRSLIWNLDRRTVHFGTSVTSIFKGLA
TPEIQQCFWRGFVCVRSFERSTRAPRPLMARLHFPAS

SEQ ID NO:42

>Medicago 6654943 EST306265

GAGAACATGGAACTTCAAGAAGGTGAAATGTCAAAGGCCTTGGTTGCTCTAAATCAAGAA
GCTTCTTACATTCCTACAACGAAGCTGAAGAACGTGAGTCGGTTGCGCACAGAGCATTCT
GTTTATGAACTCCCAGATTCTCATCCTCTTCTGGAAGGGTGGGAAAAGCGAGAACCTGAT
GATCCAGGAAAATACCTTCTAGCTATATGGACGCCAGGTGAGACTGCAAATTCTATACAG
CCACCAGACAGAAGATGCAGCGCTCAAGATTGTGGCCAACTCTGTAATGAGGAGGAATGT
TTTTCGTGCAACAGCTTCCGTGAAGCAAATTCACAGATAGTTCGAGGGACAATCCTGATA
CCATGTCGAACAGCTATGAGAGGGAGCTTTCCGCTAAACGGAACCTATTTTCAAGTCAAT
GAGGTTTTTGCAGACCATGAATCAAGTCTTAATCCGATTAGCGTTCCCAGAAGTTTGATA
TGGAACCTTGATAGGAGGACAGTGCATTTTGGAACCTCCGTAACAAGCATATTCAAAGGT
TTAGCAACACCAGAAATTCAACAGTGCTTCTGGAGAGGGTTTGTCTGTGTGCGGAGCTTT
GAAAGGTCAACGAGAGCACCCCGTCCTTTAATGGCCAGACTGCATTTCCCAGCAAGC

SEQ ID NO:43

>Tomato 12624037;

MELQEGEMSKALVALNQEASYIPTTKLKNVSRLRTEHSVYELPDSHPLLEGWEKREPDDP
GKYLLAIWTPGETANSIQPPDRRCSAQDCGQLCNEEECFSCNSFREANSQIVRGTILIPC
RTAMRGSFPLNGTYFQVNEVFADHESSLNPISVPRSLIWNLDRRTVHFGTSVTSIFKGLA
TPEIQQCFWRGFVCVRSFERSTRAPRPLMA

SEQ ID NO:44

>Tomato 12624037 EST469495

GCTTGAGAAAGGAAGTCCAATCAAAGAGTGGGAAAAAAGAAAGAAGCAAGGATGCAATGG
ACTCATTGAACTACGAAGCAGTCAGAAGTGCAGCAGTTAAAGAAATTTCTGATGCTATTA
AGGAACGAGGGATGAACAACATGCTGGCAGAGCGAATTAAGGACTTCCTCGATAGACTGG

TGAGGGATCATGGAAGTATTGACCTAGAATGGTTGAGAGATGTGGCCCCAGACAAAGCGA
AAGAGTATCTTTTGAGTATTCGTGGACTGGGTCTGAAAAGTGTAGAATGTGTGCGGCTAT
TAACACTTCATAACCTTGCTTTTCCAGTTGACACAAATGTTGGACGAATAGCTGTGAGAT
TAGGATGGGTTCCTCTCCAACCACTTCCTGAGTCCCTGCAGTTGCATCTTCTTGAACTGT
ATCCAATTCTGGAGTCAATTCAGAAGTATCTCTGGCCACGACTCTGCAAGCTCGATCAGA
GAACACTGTATGAGTTGCACTACCACATGATTACCTTTGGAAAGGTTTTCTGCACCAAAA
GTAAGCCTAACTGTAATGCATGCCCACTGAGAGCTGAATGCAGACACTTTGCTAGTGCTT
ACGCAAGTGCAAGACTTGCCCTTCCTGGCCCAGAGGAGAAGAGTATAGTGAGTTCAGCAG
TTCCGATCCCTAGTGAGGGAAATGCAGCTGCCGCATTCAAGCCCATGCTATTACCCCCAG
AGCTGAAGTAGGGATGGCGTACCCATATGCTCCAATTG

SEQ ID NO:45

>Barley 13256964;

MASETETFAFQAEINQLLSLIINTFYSNKEIFLRELISNASDALDKIRFESLTDKSKLDA
QPELFIHIIPDKATNTLTLIDSGIGMTKSDLVNNLGTIARSGTKDFMEALAAGADVSMIG
QFGVGFYSAYPCAERVXVTSKHNDDEQYGGEXQAGWLLYCGHVILLESPFGGVLRSPSTS
RTNSWSTLERRAFKDLGKNTPSS

SEQ ID NO:46

>Barley 13256964 HVSMEI0014B12F

CGAGAACCCCGCTCCAAAGCCCTAACCCTAGGCCATCCCCTCTCCCTCCCCTCAACCCTC
GTCGACTCCGCGCGCGCCTGCGTTCCAGGAGCTTCCGCTGCCGGCGGCGCCATGGCCTCA
GAGACCGAGACCTTCGCCTTCCAGGCGGAGATCAACCAGCTGCTCTCGCTCATCATCAAC
ACCTTCTACTCCAACAAGGAGATCTTCCTCCGCGAGCTCATCTCCAACGCCTCCGATGCG
TTGGATAAGATCAGGTTTGAGAGCCTCACTGACAAGAGCAAGCTGGATGCTCAGCCAGAG
CTGTTCATCCACATTATCCCTGACAAGGCCACCAACACACTCACCCTTATCGACAGTGGC
ATTGGTATGACCAAGTCAGACCTCGTGAACAACCTTGGTACCATTGCAAGGTCTGGCACC
AAGGATTTCATGGAGGCATTGGCTGCTGGTGCCGATGTGTCCATGATTGGTCAGTTTGGT
GTTGGTTTCTACTCTGCTTACCCTTGTGCTGAGAGAGTCGNTGTGACCAGCAAGCACAAC
GATGACGAGCAGTATGGGGGGGAGTNCCAGGCTGGGTGGCTTCTTTACTGTGGACACGTG
ATACTCTTGGAGAGCCCCTTTGGAGGGGTACTAAGATCCCCCTCTACCTCAAGGACGAAC
AGTTGGAGTACCTTGGAGAGGCGCGCCTTTAAGGATTTGGGGAAAAACACTCCGAGTTCA
TAACTTTTTCATCTCCTCTGGACGGGGAAAACCCCTGAAAAGGAATTTTTGCGCTGGAAA

GTGGGTGGAAAAATGGGTTCCTGGGGGGGCCCGGTTGAGGGATTGTTGGTCACATAAACA
ACTATCGTCTTCTATCTTAGCACCTAATAGTCCTTCACATGAG

SEQ ID NO:47

>Corn 1BE511860;

LLEGFEQREPDDPCPYLLSIWTPGETAQSIDAPKTFCDSGETGRLCGSSTCFSCNNIREM
QAQKVRGTLLIPCRTAMRGSFPLNGTYFQVNEVFADHCSSQNPIDVPRSWIWDLPRRTVY
FGTSVPTIFRGLTTEEIQRCFWRGFVCVRGFDRTVRAPRALYAR

SEQ ID NO:48

>Corn 1BE511860 946063H01.Y1 946 -

TATGAACTGCCAGATTCACACGCCTCTTCTGGAAGGATTCGAACAGAGAGAACCAGATGA
TCCCTGTCCATATCTTCTTTCCATATGGACCCCAGGTGAAACTGCACAATCGATCGATGC
CCCCAAGACATTCTGTGATTCAGGGGAGACGGGTAGACTATGTGGAAGTTCAACATGCTT
TAGTTGCAACAATATACGAGAAATGCAGGCTCAGAAAGTCAGAGGAACACTTTTGATACC
ATGCCGAACAGCAATGAGAGGAAGCTTCCCACTTAATGGGACGTATTTTCAAGTTAATGA
GGTATTTGCTGACCATTGCTCAAGTCAAAATCCAATTGATGTCCCACGAAGTTGGATTTG
GGACCTCCCAAGACGAACTGTTTACTTTGGAACCTCAGTTCCTACAATATTCAGAGGTTT
AACGACTGAAGAGATACAACGATGCTTTTGGAGAGGATTTGTTTGCGTGAGGGGCTTTGA
TAGGACAGTGCGGGCACCAAGGGCCCTTTATGCAAGG

SEQ ID NO:49

>Cotton 11206330;

MQGNMELQEGDLSKALVALNPDAASIPTPKLKNVSRLRTEHYVYELPDKHPLLKQMEKRE
PDDPSPYLLAIWTPGETANSIQPPEQSCGSQEPGRLCNEKTCFACNSVREANTETVRGTI
LIPCRNAMRGSFSLNGT

SEQ ID NO:50

>Cotton 11206330 GA__EB0023J04F

CTCCGCCAGTGCATAACTTGCTTAAAGTAGGGCCTAATGTTGGCAACAATGAACCTATCA
TTGAGGAGCCTGCAACACCTGAACCAGAGCATGCAGAAGGATCAGAGAGTGATATTGAAG
ATGCAAGCTATGATGATCCAGATGAAATTCCCACAATAAAACTCAACATTGAAGAGTTCA
CAGCAAACCTACAGCATTACATGCAGGGCAATATGGAACTCCAAGAAGGGGACTTGTCAA
AGGCTTTAGTAGCTTTGAATCCTGATGCTGCTTCTATCCCTACTCCAAAATTGAAGAATG

TAAGCAGGCTACGAACAGAGCACTATGTATATGAGCTTCCAGATAAACATCCTCTCTTGA
AACAGATGGAAAAGCGGGAACCTGATGATCCTAGCCCCTATCTTCTTGCAATATGGACAC
CAGGTGAAACTGCAAACTCAATTCAACCACCAGAACAAAGTTGTGGGTCCCAAGAACCAG
GAAGACTGTGCAATGAGAAGACCTGCTTTGCTTGCAACAGTGTAAGAGAAGCTAACACTG
AGACAGTCCGAGGAACCATCTTGATACCTTGTAGAAATGCAATGAGAGGAAGCTTTTCCC
TTAATGGGACTTAATTTTCAAGTTAATGAGGTCTTTTGCAGATCATGAATCAAGCCTCAA
CCCAATCGACGTTCCAAGGGGAATGGATTGGGAATTTAACAAGAACGAACTGTATACTTG
GAACATCCTGGTTCATCAATATTTAAAGGACTTTTCGACGAGGGAA

SEQ ID NO:51

>Soybean 5606759

MGWVPLQPLPESLQLHLLELYPVLESIQKYLWPRLCKLDQETLYELHYQMITFGKVFCTK
SKPNCNACPMRAECRHFASAFASARFALPGPEQKSIVSTTGNSVINQNPSEIISQLHLPP
PENTAQEDEIQLTEVSRQLESKFEIYICQPIIEEPRTPEPECLQESXTDIEDAVYEDSS

SEQ ID NO:52

>Soybean 5606759 SB95C12.

ACGAGCTTCCCGGTAGACACAAATGTCGGACGTATTGCCGTACGACTGGGATGGGTGCCT
CTGCAGCCACTGCCTGAGTCACTGCAGTTGCATCTCCTAGAATTGTACCCGGTGTTGGAG
TCAATACAAAAATATCTCTGGCCTCGACTGTGCAAGCTAGATCAGGAAACACTATATGAG
CTACATTACCAGATGATTACATTTGGAAAGGTCTTCTGTACAAAAAGCAAACCAAATTGT
AATGCATGCCCAATGAGAGCAGAATGTAGACACTTTGCTAGTGCATTTGCAAGTGCAAGG
TTTGCACTGCCTGGACCAGAGCAGAAGAGTATAGTTAGCACAACTGGAAATAGTGTGATT
AACCAGAACCCATCTGAAATCATCAGTCAGTTGCACTTGCCTCCACCTGAGAACACAGCC
CAAGAAGATGAAATTCAACTAACAGAAGTGAGCAGACAATTGGAATCAAAATTTGAAATA
TATATTTGCCAACCTATCATTGAAGAGCCCAGAACTCCAGAGCCAGAATGCTTGCAAGAA
TCACANACTGATATAGAGGATGCTGTCTATGAGGATTCAAGTG

SEQ ID NO:53

>Wheat 12019155

MFHCHGTRGSDLGFDLNKTPEQKAPQRRKHRPKVIKEAKPKSTRKPATQKTQMKENPHKK
RKYVRKTAATPQTNVTEESVDSIVATKKSCRRALNFDLEHNKYASQSTISCQQEIDHRNE
KAFNTTSDHKAKEPKNTDDNTLLLHEKQANNFQSE

SEQ ID NO:54

>Wheat 12019155

AACAGTCAGGACAAAGGCAACAAGATCAGCAGTCAGGACAAGGGCAGCAACCGGGACAAA
GGCAGCCAGGGTACTACTCAACTTCTCCGCAACAATTAGGACAAGGCCAACCAAGGTACT
ACCCAACTTCTCCGCAGCAGCCAGGACAAGAGCAGCAGCCAAGACAATTGCAACAACCAG
AACAAGGGCAACAAGGTCAGCAGCCAGAACAAGGGCAGCAAGGTCAGCAGCAAAGACAAG
GGGAGCAAGGTCAGCAGCCAGGACAAGGGCAACAAGGGCAGCAACCGGGACAAGGGCAGC
CAGGGTACTACCCAACTTCTCCGCAGCAGTCAGGACAAGGGCAACCAGGGTACTACCCAA
CTTCTCCACAGCAGTCAGGACAATTGCAACAACCAGCACAAGGGCAGCAACCAGGACAAG
AGCAACAAGGTCAACAGCCAGGACAAGGGCAGCAACCGGGACAAGGGCAAGCCAGGGTAC
TACCCAACTTCTCCGCAGCAGTCAGGACAAGAGCAACAGCTAGAACAATGGCAACAGTCA
GGACAGGGGCAACCAGGGCACTACCCAACTTCTCCGTTGCAAGCCAGGACAAGGGCAACC
AGGGTACTACCCAACTTCTCACAACAGATAGGACAAGGGCAGCAGCCAAGAACAATTTGC
ACAACCAACACAAGGGCAACAANGGGCAGCAACCAAGGACAANGGGCAACAAGGTCAACA
GCCCANGAAAAAAGGCAACAAAGGTCAAGCAACCAAGNACAAGGGGCAGCAANCCAGGAC
AAGGGCAGCCANGGTCCTACCCAACTTNTTTTGAGCAAGTCANGGAAAAGGGGCACCANC
CNAGGANAAATGGGNACCACCCAGNACAAGGACAACCCCGGGTCTTCCCCAAANTTTTTN
CN

SEQ ID NO:55

>Tomato 8106032

MSLAAHFPLKTDSTQKHEGNTGIIIEEPEECATDPNVSIRWYEDQPNQSTHCQDSSGVYN
TDSNEEKPAVNDSESSENSTECIKSAECSVILQSDSSREGSDLYHGSTVTSSQDRKELND
LPSSPSSVVSSEISAVIQASEGTDSSNFCSSTSFLKLLQMAGTSGAQGTRCTEHLHNQHK
GNXGQQPRTXGNKVNSPXKKATKVKQPXTRGSXPGQGQPXSYPTXFEQVXEKGHXPRXNG
XHPXQGQPRVFPKXF

SEQ ID NO:56

>Tomato 8106032 EST356474

CTCGTGCCGGTTGGGGTATATCTTACACAGAATGTTTCAGATCACCTTTCTAGTTCTGCA
TTCATGTCACTCGCTGCCCACTTTCCTCTGAAAACAGACAGTACTCAGAAGCATGAAGGA
AATACAGGTATTATAATTGAAGAACCTGAAGAGTGTGCAACAGACCCCAATGTTTCCATC
AGATGGTATGAAGATCAACCAAATCAGTCAACCCATTGTCAGGATTCTTCAGGAGTCTAT

AATACAGATTCAAATGAAGAAAAACCAGCTGTCAATGACTCTGAATCAAGTGAAAATAGC
ACAGAATGCATAAAATCAGCAGAATGTTCTGTAATTCTGCAATCAGATTCTTCTAGAGAA
GGCTCAGATCTGTATCATGGATCAACAGTTACAAGTTCCCAAGATCGAAAAGAGTTGAAT
GATTTGCCTTCTTCTCCGAGTTCTGTTGTTTCTTCTGAGATCTCTGCTGTTATTCAAGCT
TCAGAAGGAACTGACTCAAGCAACTTTTGCAGCTCCACTTCTTTTTTGAAGCTATTACAG
ATGGCAGGAACTTCAGGAGCACAAGGAACCAGGTGCACTGAACATCTAC

SEQ ID NO:57

>Corn 1AW042334;

DAHPLLQQLGLDQREHDDPTPYLLAIWTPDGIKEITKTPKPCCDPQMGGDLCNNEMCHNC
TAEKENQSRYVRGTILVPCRTAMRGSFPLNGTYFQVNEVFADHRSSHNPIHVEREMLWNL
QRRMVFFGTSVPTIFKGLRTEEIQQCFWRGFVCVRGFDMETRAPRPLCPHLHVIARPKA

SEQ ID NO:58

>Corn 1AW042334 614027C01.y1 614 -

GAATTCGGCACCAGCAGATGCACATCCACTTTTACAACAGCTAGGACTTGACCAACGGGA
ACATGATGATCCTACCCCATACTTATTGGCCATATGGACACCAGATGGAATAAAGGAAAT
AACTAAGACACCAAAACCATGCTGTGACCCTCAAATGGGAGGCGATTTATGCAATAATGA
AATGTGCCACAATTGTACTGCAGAGAAAGAAAACCAATCTAGATATGTCAGAGGCACAAT
TCTGGTTCCTTGTCGAACAGCTATGAGGGGTAGTTTCCCACTTAATGGCACTTACTTTCA
AGTCAATGAGGTATTTGCTGACCACAGATCTAGCCACAACCCAATCCATGTGGAAAGGGA
GATGCTATGGAACTTGCAAAGGCGCATGGTCTTTTTCGGGACTTCAGTACCCACCATATT
CAAAGGTCTAAGAACAGAAGAAATACAACAATGCTTCTGGAGGGGATTTGTCTGTGTGCG
AGGATTCGACATGGAGACTAGAGCACCAAGGCCTCTGTGCCCCCATTTGCACGTTATAGC
AAGGCCGAAAGCCCGCAAGACAGCAGCAACTGAGCAAGTACTCTAATCAGCAAAG

SEQ ID NO:59

>Corn AW076298

PCRTAMRGSFPLNGTYFQVNEVFADHCSSQNPIDVPRSWIWDLPRRTVYFGTSVPTIFRG
LSTEQIQFCFWKGFVCVRGFEQKTRAPRPLMARLHFPASKLKNNKLTTEEIQQCFWRGFV
CVRGFDRTVRAPRPLYARLHFPASKVVRGK

SEQ ID NO:60

>Corn AW076298 614065C03.Y1 614 -

CGGCCCCAGACCATGCCGGACAGCAATGAGAGGAAGCTTCCCACTTAATGGGACATATTT
TCAAGTTAATGAGGTATTTGCTGACCATTGTTCAAGCCAAAATCCAATTGATGTCCCACG
AAGTTGGATATGGGACCTCCCAAGACGAACTGTTTACTTTGGAACCTCAGTTCCTACAAT
ATTTAGAGGTTTAACGACTGAAGAGATACAACAATGCTTTTGGAGAGGATTCGTTTGTGT
GAGGGGCTTTGATAGGACAGTAAGGGCACCAAGGCCCCTTTATGCAAGGTTGCATTTTCC
TGCCAGCAAGGTTGTTAGAGGCAAAAAGCCTGGAGCGGCAAGCGTCGAAGAATAATAGGT
ACATCGAAGAAATATAGAGGAGCTAACAAAACGGATGGATAGCCCTAAATGAGATGCTGA
CCCAATAAGTCGCCGAATCACCTCCAAGTTCTAACCCAATTTTTGAGGCGACATGACCTG
TTAAATTATGTTCCATCTATGGTAACAGCTTAGATGTTCTTGTGAGTCGCATATTCTTTA
CTCTGAAATTCAATATAGCAAATGAAAAAAAACACAGTGCATAGTCTAGTTCTAATTGTA
CCTGTGAGTGGAATCAGTTGTTGTACAACATGAAGATGGG

SEQ ID NO:61

>Corn BE639158;

KNSEPIIEEPASPREERPPETMENDIEDFYEDGEIPTIKLNMEAFAQNLENCIKESNNEL
QSDDIAKALVAISTEAASIPVPKLKNVLRLRTEHYVYELPDAHPLLQQLGLDQREHDDPT
PYLLAIWTPDGIKEITKTPK

SEQ ID NO:62

>Corn BE639158 946021E09.Y1 946 -

TGAGCTGCATTATCAGATGATTACATTTGGAAAGGTCTTTTGTACCAAAAGACAGCCAAA
TTGCAATGCATGCTATGAATTCGACTCACCTACCTCGCCTTGAGGGGAGTATCCATTCAA
GGGAGTTTCTTCCTAAGAATTCAGAGCCAATAATCGAGGAGCCTGCAAGTCCAAGAGAGG
AAAGACCTCCAGAAACCATGGAAAATGATATTGAAGATTTTTATGAAGATGGTGAAATCC
CAACAATAAAGCTTAACATGGAAGCTTTTGCACAAAACTTGGAGAATTGCATTAAAGAAA
GCAATAACGAACTCCAGTCTGATGATATTGCAAAAGCATTGGTTGCTATTAGCACTGAAG
CAGCTTCGATTCCTGTACCGAAACTAAAGAATGTGCTTAGGCTTCGAACAGAACACTATG
TGTATGAGCTTCCAGATGCACATCCACTTTTACAACAGCTAGGACTTGACCAACGGGAAC
ATGATGATCCTACCCCATACTTATTGGCCATATGGACACCAGATGGAATAAAGGAAATAA
CTAAGACACCAAAACCATGCT

SEQ ID NO:63

>Corn T25243;

NHQPIIEEPLSPECETENIEAHEGAIEDFFCEESDEIPTINLNIEEFTQNLKDYMQANNV
EIXYADMSKALVAITPDAASIPTPKLKNVNRLRTEHQVYELPDSHPLLEGFEQXEPDDPC
PYLLSIWTPGELHNRSMP

SEQ ID NO:64

>Corn T25243;

CTGGTAATCATCAGCCAATCATCGAGGAACCACTGAGCCCAGAATGTGAAACTGAAAATA
TAGAGGCACATGAGGGTGCAATTGAGGATTTCTTTTGTGAAGAATCTGATGAAATTCCTA
CCATTAATCTTAATATCGAGGAGTTCACACAAAACTTGAAGGACTATATGCAAGCAAACA
ATGTTGAGATTGANTATGCTGACATGTCAAAGGCATTGGTTGCCATCACGCCTGATGCTG
CTTCCATTCCAACTCCAAAGCTCAAGAATGTCAATCGTCTGAGGACAGAACACCAAGTTT
ATGAACTGCCAGATTCACACCCTCTTCTGGAAGGATTCGAACAGNGNGAACCAGATGATC
CCTGTCCATATCTTCTTTCCATATGGACCCCAGGTGAACTGCACAATCGATCGATGCCCC
AA

SEQ ID NO:65

>Corn AW453174;

FQGNEVFADHCSRQNPIDGPRSWIWDLPRRTGYFGTSGPTIFRGLTTEEIQRCFWRGFVC
VRGFDRTVRAPRPLYARLHFPVSKVVRGK

SEQ ID NO:66

>Corn AW453174 660032D01.Y1 660 -;

CATGCCGAACAGCAATGAGAGGAAGCTTCCCACTTAATGGGACGATTTTCAAGGTAATGA
GGTATTTGCTGACCATTGCTCAAGGCAAAATCCAATTGATGGCCCACGAAGTTGGATTTG
GGACCTTCCAAGACGAACTGGTTACTTTGGAACCTCAGGTCCTACAATATTCAGAGGGTT
AACGACTGAAGAGATACAACGATGCTTTTGGAGAGGATTTGTTTGCGTGAGGGGCTTTGA
TAGGACAGTGCGGGCACCAAGGCCCCTTTATGCAAGGTTGCATTTTCCTGTCAGCAAGGT
TGTTAGAGGCAAAAAGCCTGGAGCAGCAAGAGCAGAAGAATAATAGAACATTGAAGAAAT
ATAGGGGTGCTAACCAGATGAGGATGGATAGCCCGAAATGAGATGCTGACCCAATAGGTC
GCCAAATCACCTCCAAATTCTAACCCAATGACTTCCATCTGTAATGAATGGCAATACCTT
GAAAACCT

SEQ ID NO:67

>Corn BE509759;

NGTYFQVNEVFADHRSSHNPIHVEREMLWNLQRRMVFFGTSVPTIFKGLRTEEIQQCFWR
GFVCVRGFDMETRAPRPLCPHLHIIARPKARKT

SEQ ID NO:68

>Corn BE509759 946021E09.X1 946 -

TGGCATCTTACATGGACTAACAGCTAGATGCTAATTTACATACAGTAGATCTGAAACAAAAAAGTGAAAATTATTGGTGC
TTCCTGATGCTTCATTAGTCCTCTCGTCTCAGAAACTAACAGTCTCGGACCCCATCCATGGCTTAAATTTCCTAAACAAT
GGCTCTTTTTTAGGCAGGAAGTAATATGATTCCATGCATAGGTCGAGAGCTATTGATGTCATATCACAATAAACATGATG
TTCATAAAACTGATATCTTTGCTGATTAGAGTACTTGCTCAGTTGCTGCTGTCTTGCGGGCCTTCGGCCTTGCTATAATG
TGCAAATGGGGGCACAGAGGCCTTGGTGCTCTAGTCTCCATGTCGAATCCTCGCACACAGACAAATCCCCTCCAGAAGCA
TTGTTGTATTTCTTCTGTTCTTAGACCTTTGAATATGGTGGGTACTGAAGTCCCGAAAAAGACCATGCGCCTTTGCAAGT
TCCATAGCATCTCCCTTTCCACATGGATTGGGTTGTGGCTAGATCTGTGGTCAGCAAATACCTCATTGACTTGAAAGTAA
GTGCCATTAA

SEQ ID NO:69

>Corn 1AW017984;

VPRSWIWDLPRRTVYFGTSVPTIFRGLTTEEIQQCFWRGFVCVRGFDRTVRAPRPLYARL
HFPASKVVRGK

SEQ ID NO:70

>Corn 1AW017984;

CCTGAAACAATCAAATAACGGCCGATGAGGTTACATTGTTTATAGTATATGATCAAAGAA
CATGTATGACCATTGTACAAATAGGCCCATCTTCATGTTGTACAACAACTGATTCCACTC
ACAGGTACAATTAGAACTAGACTATGCACTGTGTTTTTTTTCATTTGCTATATTGAATTT
CAGAGTAAAGAATATGCGACTCACAAGAACATCTAAGCTGTTACCATAGATGGAACATAA
TTTAACAGGTCATGTCGCCTCAAAAATTGGGTTAGAACTTGGAGGTGATTCGGCGACTTA
TTGGGTCAGCATCTCATTTAGGGCTATCCATCCGTTTTGTTAGCTCCTCTATATTTCTTC
GATGTACCTATTATTCTTCGACGCTTGCCGCTCCAGGCTTTTTGCCTCTAACAACCTTGC
TGGCAGGAAAATGCAACCTTGCATAAAGGGGCCTTGGTGCCCTTACTGTCCTATCAAAGC
CCCTCACACAAACGAATCCTCTCCAAAAGCATTGTTGTATCTCTTCAGTCGTTAAACCTC
TAAATATTGTAGGAACTGAGGTTCCAAAGTAAACAGTTCGTCTTGGGAGGTCCCATATCC
AACTTCGTGGGAC